

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2001, 20:37:40 ; Search time 33971.2 Seconds
(without alignments)
12140.628 Million cell updates/sec

Title: US-09-434-382-28
Perfect score: 26664
Sequence: 1 tatcaggtgactgaattcta.....ttcgcaagtccttttgaca 26664

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size: 8
Total number of hits satisfying chosen parameters: 2654303

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_ba3.*
- 4: gb_in1.*
- 5: gb_in2.*
- 6: gb_in3.*
- 7: gb_om.*
- 8: gb_ov.*
- 9: gb_pat1.*
- 10: gb_pat2.*
- 11: gb_ph.*
- 12: gb_pl1.*
- 13: gb_pl2.*
- 14: gb_pl3.*
- 15: gb_pl4.*
- 16: em_ba1.*
- 17: em_ba2.*
- 18: em_fun.*
- 19: em_htgo_hum.*
- 20: em_htgo_inv.*
- 21: em_htgo_rod.*
- 22: em_htg_hum1.*
- 23: em_htg_hum2.*
- 24: em_htg_hum3.*
- 25: em_htg_hum4.*
- 26: em_htg_hum5.*
- 27: em_htg_hum6.*
- 28: em_htg_hum7.*
- 29: em_htg_hum8.*
- 30: em_htg_inv1.*
- 31: em_htg_inv2.*
- 32: em_htg_other.*
- 33: em_htg_rod.*
- 34: em_hum1.*
- 35: em_hum2.*
- 36: em_hum3.*
- 37: em_hum4.*
- 38: em_hum5.*
- 39: em_hum6.*
- 40: em_hum7.*
- 41: em_in.*
- 42: em_om.*
- 43: em_or.*

- 44: em_ov.*
- 45: em_pat.*
- 46: em_ph.*
- 47: em_pl.*
- 48: em_ro.*
- 49: em_sts.*
- 50: em_sy.*
- 51: em_un.*
- 52: em_vi.*
- 53: gb_sts1.*
- 54: gb_sts2.*
- 55: gb_sts3.*
- 56: gb_sy.*
- 57: gb_un.*
- 58: gb_vil.*
- 59: gb_vil2.*
- 60: gb_vil2.*
- 61: gb_vil2.*
- 62: gb_vil2.*
- 63: gb_vil2.*
- 64: gb_vil2.*
- 65: gb_vil2.*
- 66: gb_vil2.*
- 67: gb_vil2.*
- 68: gb_vil2.*
- 69: gb_vil2.*
- 70: gb_vil2.*
- 71: gb_vil2.*
- 72: gb_vil2.*
- 73: gb_vil2.*
- 74: gb_vil2.*
- 75: gb_vil2.*
- 76: gb_vil2.*
- 77: gb_vil2.*
- 78: gb_vil2.*
- 79: gb_vil2.*
- 80: gb_vil2.*
- 81: gb_vil2.*
- 82: gb_vil2.*
- 83: gb_vil2.*
- 84: gb_vil2.*
- 85: gb_vil2.*
- 86: gb_vil2.*
- 87: gb_vil2.*
- 88: gb_vil2.*
- 89: gb_vil2.*
- 90: gb_vil2.*
- 91: gb_vil2.*
- 92: gb_vil2.*
- 93: gb_vil2.*
- 94: gb_vil2.*
- 95: gb_vil2.*
- 96: gb_vil2.*
- 97: gb_vil2.*
- 98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	26103	97.9	118788	85	AC005277	AC005277 Homo sapi
2	689	2.6	740	89	AF304371S2	AF304371 Homo sapi
3	670	2.5	721	89	AF304371S1	AF304371 Homo sapi
4	657	2.5	2908	89	AF304370	AF304370 Homo sapi
5	606	2.3	2997	91	BC001939	BC001939 Homo sapi
6	606	2.3	3006	91	BC004158	BC004158 Homo sapi
7	330	1.2	2908	89	AF308698	AF308698 Pan trogl
8	321	1.2	2976	89	AK001392	AK001392 Homo sapi

c 9	245	0.9	429	54	G60268	G60268 SHGC-130964	82	70	0.3	305583	60	AC008121	AC008121 Homo sapi							
c 10	210	0.8	220	54	G36081	G36081 SHGC-104094	83	69	0.3	60280	78	AC091132	AC091132 Homo sapi							
c 11	196	0.7	2893	89	AF308694	AF308694 Gorilla g	84	69	0.3	114411	97	HUAC002551	AC002551 Human Chr							
c 12	158	0.6	386	10	AX069570	AX069570 Sequence	c 85	69	0.3	137641	62	AC011979	AC011979 Homo sapi							
c 13	81	0.3	164242	88	AC023114	AC023114 Homo sapi	c 86	69	0.3	153620	70	AC026709	AC026709 Homo sapi							
c 14	80	0.3	105787	87	AC010458	AC010458 Homo sapi	c 87	69	0.3	154577	80	AL357046	AL357046 Homo sapi							
c 15	80	0.3	151889	60	AC008158	AC008158 Homo sapi	c 88	69	0.3	168111	69	AC025788	AC025788 Homo sapi							
c 16	80	0.3	221507	92	HS407F11	AL022329 Human DNA	c 89	69	0.3	168584	68	AC023388	AC023388 Homo sapi							
c 17	79	0.3	24703	61	AC010544	AC010544 Homo sapi	c 90	69	0.3	171901	66	AC021178	AC021178 Homo sapi							
c 18	79	0.3	121067	78	AF301505	AF301505 Homo sapi	c 91	69	0.3	177129	71	AC037481	AC037481 Homo sapi							
c 19	79	0.3	140801	63	AC015663	AC015663 Homo sapi	c 92	69	0.3	179947	61	AC009786	AC009786 Homo sapi							
c 20	79	0.3	160210	65	AC019011	AC019011 Homo sapi	c 93	69	0.3	185257	90	AL355305	AL355305 Human DNA							
c 21	79	0.3	160990	63	AC015664	AC015664 Homo sapi	c 94	69	0.3	189149	72	AC046171	AC046171 Homo sapi							
c 22	79	0.3	164353	68	AC023829	AC023829 Homo sapi	c 95	69	0.3	204685	72	AC055866	AC055866 Homo sapi							
c 23	79	0.3	165173	65	AC018844	AC018844 Homo sapi	c 96	69	0.3	222779	71	AC027808	AC027808 Homo sapi							
c 24	79	0.3	182608	65	AC018924	AC018924 Homo sapi	c 97	68	0.3	644	89	AF282036	AF282036 Homo sapi							
c 25	79	0.3	186892	75	AC078861	AC078861 Homo sapi	c 98	68	0.3	139378	92	HS64K7	AL031668 Human DNA							
c 26	78	0.3	137	54	G42927	G42927 WTAF-176-ST	c 99	68	0.3	152714	66	AC020656	AC020656 Homo sapi							
c 27	78	0.3	100272	93	HSDJ842G6	AL109657 Human DNA	100	68	0.3	152794	90	AL359085	AL359085 Human DNA							
c 28	78	0.3	141708	69	AC024993	AC024993 Homo sapi	ALIGNMENTS													
c 29	78	0.3	180052	63	AC015520	AC015520 Homo sapi														
c 30	78	0.3	213943	63	AC015465	AC015465 Homo sapi														
c 31	77	0.3	79666	89	AL136969	AL136969 Human DNA														
c 32	77	0.3	149496	72	AC053493	AC053493 Homo sapi														
c 33	77	0.3	175197	74	AC073283	AC073283 Homo sapi														
c 34	77	0.3	176626	88	AC084754	AC084754 Homo sapi														
c 35	77	0.3	222548	67	AC022202	AC022202 Homo sapi														
c 36	76	0.3	131684	89	AL137849	AL137849 Human DNA														
c 37	76	0.3	134601	81	AL450106	AL450106 Homo sapi														
c 38	76	0.3	149386	71	AC040997	AC040997 Homo sapi	ORGANISM													
c 39	76	0.3	165605	81	AL391994	AL391994 Homo sapi														
c 40	76	0.3	170195	80	AL357556	AL357556 Homo sapi														
c 41	75	0.3	110414	92	HS281H8	AL031133 Human DNA														
c 42	75	0.3	120630	92	HS28F12	AL031657 Human DNA														
c 43	75	0.3	142515	86	AC008556	AC008556 Homo sapi														
c 44	75	0.3	155952	85	AC004534	AC004534 Homo sapi														
c 45	74	0.3	91835	92	HS26H23	Z84467 Human DNA s														
c 46	74	0.3	118150	93	HSJ800C24	AL121593 Human DNA														
c 47	74	0.3	125937	92	HS69M21	AL031735 Human DNA								REFERENCE						
c 48	74	0.3	135252	81	AL512423	AL512423 Homo sapi														
c 49	74	0.3	155190	70	AC026323	AC026323 Homo sapi														
c 50	74	0.3	159057	68	AC023995	AC023995 Homo sapi														
c 51	74	0.3	170795	66	AC021346	AC021346 Homo sapi														
c 52	74	0.3	176731	67	AC022735	AC022735 Homo sapi														
c 53	74	0.3	200484	69	AC025002	AC025002 Homo sapi														
c 54	73	0.3	85543	87	AC010553	AC010553 Homo sapi														
c 55	73	0.3	157784	90	AL360232	AL360232 Human DNA														
c 56	72	0.3	163884	63	AC013245	AC013245 Homo sapi	TITLE													
c 57	72	0.3	170970	79	AL354707	AL354707 Homo sapi														
c 58	72	0.3	171160	89	AL138966	AL138966 Human DNA														
c 59	72	0.3	171173	81	AL445592	AL445592 Homo sapi														
c 60	72	0.3	193830	82	AL590108	AL590108 Homo sapi														
c 61	72	0.3	208248	61	AC010900	AC010900 Homo sapi														
c 62	71	0.3	57304	85	AC004802	AC004802 Homo sapi														
c 63	71	0.3	65916	76	AC083783	AC083783 Homo sapi														
c 64	71	0.3	140368	74	AC073254	AC073254 Homo sapi														
c 65	71	0.3	153430	72	AC064852	AC064852 Homo sapi								JOURNAL						
c 66	71	0.3	188679	67	AC023121	AC023121 Homo sapi														
c 67	71	0.3	192064	84	CNS01RHC	AL161669 Homo sapi														
c 68	71	0.3	193816	77	AC087691	AC087691 Homo sapi														
c 69	71	0.3	198582	85	AC005291	AC005291 Homo sapi														
c 70	71	0.3	203748	74	AC069250	AC069250 Homo sapi														
c 71	70	0.3	94205	86	AC007639	AC007639 Homo sapi														
c 72	70	0.3	130526	89	AL133282	AL133282 Human DNA														
c 73	70	0.3	147990	69	AC026144	AC026144 Homo sapi														
c 74	70	0.3	162472	83	CNS01DUW	AL133371 Homo sapi	REFERENCE													
c 75	70	0.3	165464	67	AC022076	AC022076 Homo sapi														
c 76	70	0.3	169533	63	AC013244	AC013244 Homo sapi														
c 77	70	0.3	171841	60	AC007940	AC007940 Homo sapi														
c 78	70	0.3	179743	79	AL162400	AL162400 Homo sapi														
c 79	70	0.3	180462	75	AC074011	AC074011 Homo sapi														
c 80	70	0.3	193159	61	AC009727	AC009727 Homo sapi														
c 81	70	0.3	194405	71	AC032004	AC032004 Homo sapi														

TITLE Direct Submission
JOURNAL Submitted (23-JUL-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jul 23, 1998 this sequence version replaced gi:3335015.
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Only the first 118.8 kilobases of this clone are being submitted.
The remainder overlaps accession number AC005274 (WICGR project L350).

FEATURES	source	Location/Qualifiers	repeat_region
		1..118788	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="hRPK.597_M_12"	
		/clone_lib="RPCI-11 human BAC library"	
		/map="17"	
		/chromosome="17"	
		795..884	
		/rpt_family="MIR"	
		complement(932..1000)	
		/rpt_family="MIR"	
		2044..2076	
		/rpt_family="AT-rich"	
		2237..2713	
		/rpt_family="MER34"	
		2718..3477	
		/rpt_family="L2"	
		complement(3478..3821)	
		/rpt_family="MLT1A"	
		3822..4007	
		/rpt_family="L2"	
		complement(4015..4281)	
		/rpt_family="MER89"	
		complement(4282..4321)	
		/rpt_family="(TAAA)n"	
		complement(4328..4608)	
		/rpt_family="AluSx"	
		complement(4609..4895)	
		/rpt_family="MER89"	
		complement(4896..5186)	
		/rpt_family="AluSp"	
		5219..5473	
		/rpt_family="L2"	
		complement(5481..5740)	
		/rpt_family="MER34"	
		complement(5963..6275)	
		/rpt_family="AluSp"	
		complement(6411..6494)	
		/rpt_family="MER3A"	
		6830..7013	
		/rpt_family="MLT1C"	
		7014..7318	
		/rpt_family="AluYb8"	
		7319..7581	
		/rpt_family="MLT1C"	
		7721..8032	
		/rpt_family="AluYb"	
		8039..8333	
		/rpt_family="AluSg"	
		9431..9460	
		/rpt_family="AT-rich"	
		9662..9829	
		/rpt_family="MIR"	
		complement(11302..11728)	
		/rpt_family="MER39"	
		complement(11824..11884)	
		/rpt_family="LTR16C"	
		complement(11928..12108)	
		/rpt_family="LTR16C"	
		12453..12725	
		/rpt_family="AluYb"	
		/rpt_family="MSTB"	
		14118..14153	
		/rpt_family="(CA)n"	
		complement(15320..15620)	
		/rpt_family="AluJo"	
		15683..15759	
		/rpt_family="(CA)n"	
		complement(15838..15893)	
		/rpt_family="MER91A"	
		complement(15915..16080)	
		/rpt_family="MER91A"	
		complement(16474..16544)	
		/rpt_family="L2"	
		17907..17920	
		/rpt_family="LIME"	
		18271..18311	
		/rpt_family="L2"	
		complement(18368..18529)	
		/rpt_family="AluYb"	
		complement(18530..18560)	
		/rpt_family="(CA)n"	
		complement(18561..18681)	
		/rpt_family="AluYb"	
		18760..19065	
		/rpt_family="AluSg"	
		complement(19638..19896)	
		/rpt_family="MIR"	
		complement(19971..20260)	
		/rpt_family="LIME3A"	
		20263..20552	
		/rpt_family="MER58B"	
		complement(20553..20816)	
		/rpt_family="LIME3A"	
		complement(20821..21069)	
		/rpt_family="LIM4"	
		complement(21071..21100)	
		/rpt_family="(TAAA)n"	
		complement(21101..21338)	
		/rpt_family="AluSx"	
		complement(21348..21598)	
		/rpt_family="LIM4"	
		21799..22104	
		/rpt_family="AluJo"	
		22151..22362	
		/rpt_family="MSTA"	
		23978..24287	
		/rpt_family="AluSg"	
		24852..25199	
		/rpt_family="L2"	
		complement(25697..25985)	
		/rpt_family="MLT1D"	
		25816..25895	
		/note="Single-stranded coverage."	
		complement(25994..26099)	
		/rpt_family="purine-rich"	
		complement(26122..26291)	
		/rpt_family="MLT1D"	
		26312..26337	
		/rpt_family="(CAAAA)n"	
		26456..26480	
		/rpt_family="(GGGAA)n"	
		complement(26576..26781)	
		/rpt_family="MER6B"	
		complement(26912..26990)	
		/rpt_family="MER21B"	
		26991..27290	
		/rpt_family="AluSx"	
		complement(27291..27975)	
		/rpt_family="MER21B"	
		28034..28197	
		/rpt_family="LTR28"	

[illegible]

Db 79633 TTGGGGGCTTAAGTGGTGAGTATATTCCTTTGTCAGTGTACAGAGCTGGTGGGAAGTCTCTG 79692
Qy 2041 ggaatttaaccggctttaccatttttccaaagctcggggtggcgagctactttttttttt 2100
Db 79693 GGATTTTAACGGCTTTACCATTTTCCAAAGTCTGGGGTGGCAGCTACTTTTTTTTTT 79752
Qy 2101 ttttttttttgcagtcggtgagtccttggctcaactgcaacctttgctcttgggctcag 2160
Db 79753 TTTTTTTTTTGTCAAGTGGCGTGATCTTGGCTCACTGCAACCTTTGCCTTCTGGGCTCAG 79812
Qy 2161 gtagtccctcaacctcaagctcccaaatagctgggacacacacgctgccccatcacact 2220
Db 79813 GTGATCCCTCACTCAGCTCCCAATAGCTGGGACACAGTGTGCCCATCACACCT 79872
Qy 2221 ggcataatttttgcagtcggttagcagcggttttgcctatgttgcacagctggtc 2280
Db 79873 GCGTAATTTTTTGTATGTTTGTAGCAGCGGGTTTTGCTATGTTGCCACGGCTGGTC 79932
Qy 2281 tcaaacctctgcgatacctcctgtctcggcctccagagtgctgggattacaggtcagc 2340
Db 79933 TCAAACTTCTGCGATCCTCTGCTCGGGCTCCAGAGTGTGGGATTACAGCATGAGC 79992
Qy 2341 caccgcacctggcctggaatttcttttataccagccagccagcgacgacagcagcatt 2400
Db 79993 CACCCGACCTGGCCTGGAAATCTCTTTATACGAGCCAGTCAAGCAGCAGAGCAAT 80052
Qy 2401 aaaagctgactcagggaacagagattttaatatgataccacctcttaagttacacat 2460
Db 80053 AAAAGCTGTACTCAGGGAACAGATTTTAATATGATACCACTCTTAAGTGTACCAT 80112
Qy 2461 ccacttagttcttcggttcggggagacagagatttggcgagtaaacctggaggtcagc 2520
Db 80113 CCACTTAGTTTCTGCGTTCGGGGACACAGATTTGTGGCAGTAACTGGAGAGTCTAGC 80172
Qy 2521 agtgggtatacagttaatatgtttaccgcagacgccaattggcacattggcgacacaa 2580
Db 80173 AGTGGTGATTACAGTTAAATGTATTACCAGCAGCGCAATTGGCACATTTGGCAGCCACAA 80232
Qy 2581 cataccactgtccagattaccctgctcatttatgtctctatcaaccggaagctcagattgt 2640
Db 80233 CATACCCACTGTCCAGATTACCTGTCAATTAATGTCTATCAACCAGGAGGTGAGGATTGT 80292
Qy 2641 gttgcagccaaattgtgtggcttggcgatggacgggaaaggagtgaaggttagacca 2700
Db 80293 GTTGACGCCAAATTGTGTGGCTTGTGGCATGGACCGGAGGAGTGAAGTTAGACCA 80352
Qy 2701 gctccctctcaggctagactagggtgaggaacttaggtgcagccctcacttgc 2760
Db 80353 GTCCTCCCTTCTCAGGGCTGAGACTAGGGTGAAGCACTTAGGGTGGCAGGCCCTTCACTTGC 80412
Qy 2761 atgattccttaccattttgcacactgggtgcttgcctgtgcttaccctagtgacagctcagc 2820
Db 80413 ATGATTCTCTACATTTTGCACACTGGGTGCCTTGCTGTCTACCCCTAGTGACAGCTCAGC 80472
Qy 2821 ccattctagaggcattttaagaataattgtgtctgtttacacctctagctggcaccatt 2880
Db 80473 CCATTCTAGAGGCATTTAAGAAATATTGGTGTCTGTACACCTCTAGTGGCATCACTT 80532
Qy 2881 ctgctctgtacatcttccctgggtgttacttccaaagctggaaggtgagatgtagataaa 2940
Db 80533 CTGCTCTGTACACTTCTCCCTGGTTGTACTTCCAAAGCTGGAAGGTGGAGATGTAGATAAA 80592
Qy 2941 tagttgattataggggtgctcctcctgttagtgacacaggttcaaatgtagagaga 3000
Db 80593 TAGTTGGATTAGTACGGGGTGTCTCTCTGTGTAGTGACACAGGTCAAATGTATGAGAGA 80652
Qy 3001 tctgattttatgcactcttttaggaatgatcttacttactttaaggaacacgggcttccaa 3060
Db 80653 TCTGATTTTATGCATCTTTTAGGAATGATCTTACTTTAAGGAAACCGGGCTTCCAA 80712
Qy 3061 agtgttaactttctgacctccacaactgggtagcttcttccctgacacatctttcacaagc 3120
Db 80713 AGTGTGTACTTTCTGGACCTCCACAACCTGGTGAGTCTTTTCCCTGACACATCTTTCAAAAGC 80772

Qy 3121 aaatcttcccttttgaataatcagtaacaagaatttttccctttttgcaaatcagctcttctgc 3180
Db 80773 AATCTTTCCTTTTGTAAATATCAGTAAACAAGATTTTCCCTTTTGTGCAAAATCAGTCTCTGC 80832
Qy 3181 cctccagagataccttgggttgaacagcttcccttttcaagttaaaaagacttggagtc 3240
Db 80833 CCTCCAGAGATACCTGGTGGTTGAAAACGCTCCCTTTCAAGTTAAAAAGACTTGAGTTC 80892
Qy 3241 tgataaactatgtgaccttgatcaagttaactttaccctttctgagcttttagttttattcatc 3300
Db 80893 TGATTAAACTATGTGACCTTTGATCAAGTTACTTTACCTTTCTGAGCTTTTASTTTATTTCAATC 80952
Qy 3301 tataagatacactatcacgttttcataagagttgttaaagataaaatgacgtagcagcacata 3360
Db 80953 TATAAGATGACATATCACGTTTCATAGAGTGTGTAAAGATTAATAGACGTAGACGACACATA 81012
Qy 3361 taaagcactaaatcacctttattagatatattgttggcaccacaagtaggcacacaagaagg 3420
Db 81013 TAAAGCACATAATCACTTTATTAGATATATGTTTGGCACCAGTAGGCACACAAGAAAGG 81072
Qy 3421 gagcttttggtttttattcaataaatttctgacatcttaccctttctagctccagcttat 3480
Db 81073 GCAGCTTTTGTGTTTATTCATAAATTTCTGACATCTTCTTACCTTTCACTGACGCTTAT 81132
Qy 3481 tacactcttgagaagcggtgtgtgtgttgaataaacaagttcattttccagctcctta 3540
Db 81133 TACACTCTTGTAGAAGCGGTGTGTGTGTGAATATATACAGTTCATTTTCCAGTCCCTTA 81192
Qy 3541 agaaagaactcaccagacactgttaaagcttctcccaaaaataacgtttgaataccatcca 3600
Db 81193 AGAAGAAGTCAACCAAGACCTGTTAAGTCTTTCCCAAAAATAACGTTTGAAATCCATCCA 81252
Qy 3601 ttgtctcttattgagccttcccttatttctgttttctatgacctgtaaacatacaatagcc 3660
Db 81253 TTTGTCTCTTATTGAGGCCCTTCCCTTATTCTGTGTTTCTATGCCCTGTAAACTACAATAGCC 81312
Qy 3661 tcccatattcatctcgccctcctgttaaccatctgccacacagcagcagcagagagtcac 3720
Db 81313 TCCCATATTCTATCTCGCCCTCTCTGTAAATCCATCTGCCACACAGACGCCAGAGAGTCAAC 81372
Qy 3721 ttcaagacagaaaaagtagtgtctacttggccaccttaaagcccttcatgggtccccatt 3780
Db 81373 TTTCAAGACAGAAAAAGTAGTGTGTCACTTCCACCCCTAAAGCCCTTCATGGGTCCCAATT 81432
Qy 3781 gcaatacaatacaaacacttgatatgacctacaagtcctgttaggcccgcgcgtacacc 3840
Db 81433 GCAATAACAATCAAAACACCTTGATATGGCTTACAAAGTCTGTAGGCCCGCGCGCTACCC 81492
Qy 3841 acacttccatctgtaccatcgctgaactgacgtgcatggctgactcttatgtccctc 3900
Db 81493 ACATTCCCATCTGTACCCATCGCTGAACITGCAGCTGCATGGGCTGACTCTTATGTCCCCTC 81552
Qy 3901 taactccctggcacttccaggactttcgcccttccgcgggttccctctcgctcttctaatt 3960
Db 81553 TAACCTCCCTGGCACCTTCAGGACTTTTCGCCCTTCCCGGGTTCCTCTGCCCTCTTAAT 81612
Qy 3961 tgcctcctatatgtttactgaaccttcaggctcaggtcagctagaggggtcatttaactccagaac 4020
Db 81613 TGTGCTCTATATTGTTTACTTGAACCTTCAGGGCTCAGCTAGAGGGTCAATTACTTCCAGAAC 81672
Qy 4021 tgcctctctctctagacaagttggatcccagctctctgtatttttccatttttcccttga 4080
Db 81673 TGCCTCTCTCTCTCTAGACAAGTTGGATCCAGCCCTCTGTATTTTTCATTTTTCCTTGCA 81732
Qy 4081 gagcacttagcataaagcactaaagctgttctgttatcggttttccctttgtctcctcc 4140
Db 81733 GAGCACTTAGCATAAATGCCACTTAAGCTGTTCTGTATATCGTGTCTCTTTTGTCTCTCC 81792
Qy 4141 actggcctgattaggaagcctccatctcttttccctgctatatccttggcactgat 4200
Db 81793 ACTGGCTGATTAGACGAAGGCTCCATCTCTTTTCCCTGCTATATCTTGGCATCTGAT 81852

QY 4201 ataattgataactcagtaaaatatttggtaataaaatgatgttccaataatatttactaagctttg 4260
Db 81853 ATAATGGATACTCAGTAAATATTGTAATAAATCATGTTCAAAATATTATTACTAAGCTTGG 81912
QY 4261 ttatagtgaatacctattgtaaaccttttaataacttgaatagtgtgttctacat 4320
Db 81913 TTTATAGTTGATACCTATTGGTAACCTTTTAAATACATGTAATAGTTGCTGTCTACAT 81972
QY 4321 ttgttcaaccataactgctcatcttctgttttctcattaggaaaaaacacctcgaaagaat 4380
Db 81973 TTGTTCAACCATTAACGTCTCATTTCTTTGTTTTTCATTAGAAAAATACCTCGAAGCAAT 82032
QY 4381 caaaatatttctgggtccattgaaaggaaatagaactgggttacgtcttctgtctgactca 4440
Db 82033 CAAAATATTTCTGGTCCATTGAAGGAATAGAACTGGGTACGCTTTTGTGCTGTGACTCA 82092
QY 4441 tctctgctatttctaaacttatatatgcccctgacctctcaaataggaaatccattaaaaac 4500
Db 82093 TCCTCTGCTATTTCTAACTTATATATGCCCTGACCTCTCAAAATAGAAATCCATTAAAAAC 82152
QY 4501 atcaaatcaaacctcaaaatcaaatgcttcataccacacagagattttttttttttttt 4560
Db 82153 ATCAAAATCAAAACCTCAAAATCAAAATGCTTCATACACACAGAGATTTTTTTTTTTTTT 82212
QY 4561 tttttggatagagtctgttcttaccaggctggagtcagtgcatgatctcgctc 4620
Db 82213 TTTTTTGGATAGAGCTTGTCTTTTACCAGGCTGGAGTGCAGTGCATGATCTCGGCTC 82272
QY 4621 actgaaacctccacctctctgggttcaagccattctccacctcagcctctctgtagcta 4680
Db 82273 ACTGCAACCTTCCACCTCTCGGTGTTCAAGCCATTTCTCCACCTCAGCCTCTCGAGTAGCTA 82332
QY 4681 ggactcagcgcatgccatccacgctcagctcaattttttgtatttttagtagagacggg 4740
Db 82333 GGACTACAGCGCATGCCATCAGCTCAGCTAAATTTTTTGTATTTTGTAGTAGAGCGGG 82392
QY 4741 ttccacctgttggcaggatagctcgtatctcttgacctgtgatctgcccgcctcagc 4800
Db 82393 TTTTACCATGTTGGCCAGGATAGTCTCGATCTCTTGACCTTGTGATCTGCCCGCCTCAGC 82452
QY 4801 ctcccaaatgaactccactgtggttggagatgggatttctaaatagtgacatttctgt 4860
Db 82453 CTCCCAAAATGAGCTACCATGTGGCTGGAGATGGGATTTCTAAATAGTGACATTTTCTGT 82512
QY 4861 gttccacacctcgtctaaaaataggccaggtcgcgaggagtgattgaacagctgat 4920
Db 82513 GTTCCCACTCATGCTGTAAAAATAGGGCCAGGTGCGCAGGAGTGATTTGAACAGCTGAT 82572
QY 4921 gctgctgtgtacatgctgtgtggcatttcccatccagacggcgagggctcctgacctcag 4980
Db 82573 GCCTGCTGTGTACATGCTGTGTGGCATTTCTCCATCCAGACGCGAGGCTTCCTGCCTCAG 82632
QY 4981 ttccagaggtgcttctcgctgttgaattgcttggattggggcgggggtgacaagggtt 5040
Db 82633 TTCCAGAGTGTCTCTCGTGTGAGTTGCTTTGAGTTGGGGCGGGGGTGTACAAGGGTT 82692
QY 5041 ccttagaggttttgtgccaactttgtacattgaacacagctcccagctgcgcagggggg 5100
Db 82693 CCTAGAGGTTTTGTGGCCNACTTTGTACATTTGAACGCGAGCTCCAGCTCGCAGGGGG 82752
QY 5101 cttacagcctcttgatgggaaggcctcactgagatgctagtagggctcttgcctgg 5160
Db 82753 CTTACAGCCTCTTGATGGGAAGAGGCTCCTACTGAGGATGCTAGTAGGGCTCTTTGCTCTGG 82812
QY 5161 cactggtgtatctgtgcttgttaataactcctcttttatagaacacactaataactttgt 5220
Db 82813 CACTGTGTGTATCTGTGGCTTTTAAATACTCTCTTTTATAGAAAACACTAATACTTTGT 82872
QY 5221 ttcaaaatatacatcagctcttctgtttgctgatgatagttccctggcttccactattct 5280
Db 82873 TTCAAAATATACATCAGCTCTTCTGGTTTCCGATGATAGGTTCCCTTGGCTTCACTATTCT 82932
QY 5281 gtttggtaacttgggtctctctagccagtgctgtgatgagctgacagcactccacttgc 5340

Db 82933 GTTTGTTAACTTGGTCTCTGAAAGTTGAGTACTAGTTTCTGTCTTCAATTTTTTAACG 82992
QY 5341 gatagtcaccaagattataatgtcttttcatctggctgtagtaaaatataaaatggctgac 5400
Db 82993 GATAGTCACCAAGATATATATGTCTTTTCATCTGGCTGTAGTAANAATATAAATGGCTGAC 83052
QY 5401 caaaatcacacttttatttatttctctaaaaatggtaactctctcttagaaagctggttttcg 5460
Db 83053 CAAAATACACTTTTATTTTCTTAAAAATGGTAATCTCCTTAGAAAGTCTGGTTTTTCG 83112
QY 5461 tgtcagattccccaccobaattctgaggcaattcagttgtcgtgggttggtgatacctgaag 5520
Db 83113 TGTCAGATTTCCCACTAATTTCTGAGGCAATTCAGTTGCTGCTGCTGGTGGTGCATCCTGAAG 83172
QY 5521 ttactctccacacacatcttcactaatgaatcaccttctgtgtgtgtgttttcttcta 5580
Db 83173 TTACTCTTCCACACATCTTCACTAATGCAATCATTGCTGTGTGTGTGTCTTTTGTGTA 83232
QY 5581 gctgtcggggccacctctgcccagaaatcacgagatgaacacacatgacagtttaccagatc 5640
Db 83233 GCTGTGGGGCCCACTCTGCCCCAGAAATACAGGATGAACCATGACAGTTTACCAGATC 83292
QY 5641 cccatatcacagtgatgataaagccagggttcccagaggagggtgtacgtccctgagtaa 5700
Db 83293 CCCATACACAGTCAGTATGAAGCCAGGTTTCCAGAGGAGGGTGTACGTCCTCGAGTAA 83352
QY 5701 agaaacatgatatgaataatagaacactgaacacttgcgtgtgggcacctgttttggttct 5760
Db 83353 AGAAACATGGAATGAATAAGAACTGAACACTGTCTGTGGGCACCTGTGTGTGTCT 83412
QY 5761 gagcatgattagaaaaattagttgaggaaatgaagatatggtcctcctgcctggctataaa 5820
Db 83413 GACCATGATTAGAAAATTTAGTTGAGGAATCAAGATATGCTCCTGCCCTGGCTTATAA 83472
QY 5821 cttacgagtgctgacttatgcctaattgatgtgatttatgcttcttggaaattagataatc 5880
Db 83473 CTTTACGAGTCTGCTGACTTATGCTTAATGATGATGATTGCTTTTGGAAATATTAGATAATC 83532
QY 5881 aagcactgttgtaataatagattgcattcaagtttgcacattcattcgttggagggtttttt 5940
Db 83533 AAGCACTGTGTGTAATAGATTGCATTTCAAGTTTGACATTCATTGCTTGGAGGTTTTTT 83592
QY 5941 cccacagggttaataaccctcttttgcacagcatcatgaagaggtttgcacagatagat 6000
Db 83593 CCCACAGGCGTAATACCTCTTTTGTATCAGACCATCATGAAGAGGTTTGCACAGATAGAT 83652
QY 6001 ttttttaaaataaatgattacagcaacacctaaaaagtggttgttggggggttagaagct 6060
Db 83653 TTTTTTAAATAAATAATGATTACAGCAACCTAAAGAAAGTGTGTGTGGGGTTTGAAGCT 83712
QY 6061 cctgcaaatccgaagatatcaggccagatgatgtggtcttagcttaggaaaaaggttag 6120
Db 83713 CTTGCAAAATTTCCGAAGTATCAGGGCCAGATGATGTGTGTCTTAGGTAAGAGTTAG 83772
QY 6121 tctgtccttgaacttggctaaagacattcatctcgtgttttacttacaatgtaagaagag 6180
Db 83773 TCTGTCTTGAACCTTGGCTAAAGACATTCATGCTGTTTTTACTTTACATGTGAGAGAG 83832
QY 6181 taccagcagtaggggtatttcttctttagtactaactaatgtgagtcttactaagttagt 6240
Db 83833 TACCAAGCAGTAGGGGTATTTCCTTGTAGTACTAATAATGTGATGCTTTACTTAAGTAGT 83892
QY 6241 gctgatgggtgacagaccagagcacccagcaaaagggcagagaagctccagaacctggcag 6300
Db 83893 GCTGATGGGTGACAGACACAGAGCACCCAGCAAGGCCAGAGAAGTCCAGAACCTGGCGAG 83952
QY 6301 gagatgaggtctacactgactgaagcagaagcagcagggagggaggaatgtgccgga 6360
Db 83953 GAGATGAGGCTTACACTGACTGCTGAAGCAGCAAGACGACGAGGAGGAGGAATGTCCGGA 84012
QY 6361 gcaatggcaaaagtgtcctctagccagtgctgtgatgagctgacagcactccacttgc 6420

Db 84013 GCAATGGCACAAAGTGCTCTAGGCGAGTGCTGTGATGAGCTGTATCAGCACTCCCAATTGGC 84072
Qy 6421 tggcttgcctcctcgtcagatgccttctcctcagccctcgtcgtagcaacccccca 6480
Db 84073 TGGCTTGCTCCTCCTGCTCAGATGCCCTCTCTCACCCTGACCCCTGCTGTAGCCACCCCA 84132
Qy 6481 gctcagtgatccacacctgtttgttgcatttccagcaccctgttctcctcgtccatgg 6540
Db 84133 GCGTGAGTGCATCCACTGTTGTTGTCATTTCCAGCACCCTGTCTTCGCTCCATGG 84192
Qy 6541 catgacagtaactttcatatgtatgttgcgtgatcgtatgttaacatgctcagtttg 6600
Db 84193 CATGTGACAGTAACTTTCATATGATTTTGCCTGATCGATGTTAAACATGCTCAGTTTG 84252
Qy 6601 ccgatacgttttttcagtgctcagcgccctcagtgagtgaaactacgttcaattctcg 6660
Db 84253 CCGATCACTGTTTTTTCAGTGTCAGCGGGCCCTCAGTGAGTGAACCTTACGTTCACTTCG 84312
Qy 6661 ttgcagctgtctttagcttcttagcagcgaattttttcccttgcattcgtcagcctta 6720
Db 84313 TTGCAGCTGTGTTTACGTTCTTAGCAGCGAATTTTTCCTTGTATCTTGAGCCTTA 84372
Qy 6721 actaaatgaataatgaggtcctcttgcagataggttaacctttgggtctatgtttag 6780
Db 84373 ACTAAATGTAATAGAGCTCCTCTTGAGATAGGTACCCCTTGGGTCTATGTTGTAG 84432
Qy 6781 cggagtgatgataataaagcatgcttacaacccacatgctgtttagataacagtt 6840
Db 84433 CGGAGTGATGATAATAAAGCATGCTTACAACCCACATGCTGTTTAGATAACAGTT 84492
Qy 6841 gttgagttggttactgtggtcgaggtcgtgagcgaagcagaaacataaataatagac 6900
Db 84493 GTTGAGTTGGTACTGTGGCCGAGGCTGTGAGCTAAGCAGAAACATTAACATTAATAGGAC 84552
Qy 6901 ataggtgcagcccaagaccaggttaggaagttaactaactagttatttccactgtatag 6960
Db 84553 ATAGGTGCAGCCAGAACACAGTAGGAGTAACTAACTAGTTATTTCCTACTGTATAG 84612
Qy 6961 taaaaggtgctgatattaattggcgttctggcattcccatgatgaacgtctggcctt 7020
Db 84613 TAAAGGTGCTGATTAATTTGGCGTCTGGCATTCOCATGTATGAACGCTGGGCCCT 84672
Qy 7021 ggcgtcagctcactgtgcagtggttaatttgggtatctgactgaccaggtgaac 7080
Db 84673 GCCTGTCACTCACTGTGTCAGTGTGTAAATTTGGTGTATCTGTACTGACCAAGTGAAC 84732
Qy 7081 agaggggggaaagcacaacccaatggcagagctccagaaagccctcagcaggtccagtc 7140
Db 84733 AGAGGGGGAAAGCACCACCATGGCAGAGTCCAGAAAGGCCCTCAGCAGGCTCAGTC 84792
Qy 7141 cagagcgatcttcagactccagtygaatgaaatgagccacacacttccacatggaata 7200
Db 84793 CAGAGCGATCTTCAGACTCCAGTCCGAATGAAATGAGCCACACCTTCCACATGGTAATA 84852
Qy 7201 gtataacaaacacagacagcagaaggttcggttttcttaattctcgtcctgtgaatg 7260
Db 84853 GTATAACAAACACAGACGACGAGAAAGGCTTGGCTTTCTTAATCTCTGCTGTGAATG 84912
Qy 7261 ctgtagagatcattatttgaagaagccaggtgtgtaaacagatccttctcctgggc 7320
Db 84913 CTTGTAGAGATCATATTGTGAAGAACCCAGGTGTGTAAACAGATCCTTCTCTCGGC 84972
Qy 7321 ttactataacttgcccgcttggggaaatgagaaggggttgggtgtaaaagtggcagcctgca 7380
Db 84973 TTACTATAACTTGGCCCGCTTGGGGAAATGAGAAGGGTTGTTGTAAGGTGGCGCCTGCA 85032
Qy 7381 acttaataatgaccagtcacagtttggccacccaggtcgtggtaggcccacaaactg 7440
Db 85033 ACTTAAATGAACAGTCCACAGTTTGGCCACCCAGGGTGTGGGTAGGCCCAAACTG 85092
Qy 7441 tgtctgttttccagagaggaacagggcctgacaaacggattcatttcttatttcat 7500
Db 85093 TGTCTGTTTTCCAGAGGAGACAGGGCTTGACAAACGGATTCAATTTGTTATTTTTCAT 85152

Qy 7501 taatgtaacatttatgcataattttccattaatgtggaaactataactgctgaagccaatga 7560
Db 85153 TAATGTAACTTTATGCAAAATTTTCCATTAATGTGAAACATATAACTGCTAAGCCAAATGA 85212
Qy 7561 gacagtcataatcagtgagaggtcctgcagctctccagaatgacagccactgggaaacg 7620
Db 85213 GACAGTCAAAATCAGTGAGAGGCTCTGCACGCTCTCCAGAATGACAGCCACTGGGAAAGC 85272
Qy 7621 gagttaaaagtcgaagatgagatgtagctcagggagtcagcgcttcgggagtttggttgt 7680
Db 85273 GAGTTAAAGTCCAAAGATGAGATGATGATCAGAGTCAAGCGCTTCGGGAGTTGTGTGT 85332
Qy 7681 ccttaacagaaggttcagcgttggcaaaagctcggcagctcctcttctcctcaggtctt 7740
Db 85333 CCTTAACAGAAGGTGAGCGTTGGCAAGCTCGGACGCTCCTCTCTCTCTCTGAGGTCTT 85392
Qy 7741 gctcagtgactgagaacagggctgacccctatgtcgtccttggtttggatggcacgggt 7800
Db 85393 GTCTAGTGACTGAGAACAGAGGCTGACCCCTATGTGCTGCTCTTGTGGATGGCACCGGT 85452
Qy 7801 aaagactgaacccagcatttctcctgcagggccttgaacatttctgttatttctcatatt 7860
Db 85453 AAAGACTGACACCAAGCATTTTCTGCAAGGCTTTGAACTTTTGTGTATTTCATATAT 85512
Qy 7861 atatgttataaagcacattacaatatatttctcgtctctcagtcctcagtcctagtgaa 7920
Db 85513 ATATGTGTTATAAGCACATTAACAATATATTTTCTCTCTCTCTCTCAGTCTTAGTGAA 85572
Qy 7921 atgtcatttataaaaaaatttccactgcccatttctaaagttttctgctgagagtttctg 7980
Db 85573 ATGTGCTATTTAAAAAAATTTCACTTGCCTTCTTAAAGTTTCTGCTGAGAGTTTGT 85632
Qy 7981 gtttttcaattacgcaaacacacatctccacataagtagggaaaaaagtccttcttagtat 8040
Db 85633 GTTTTTCATTTACGCAACACATCTCCACATAAGTAGGAGGAAAAAGTCTTCTTGAGTAT 85692
Qy 8041 attagtcctcagccttctgtattgggacagtagcgtccatttaatttcttctgagtgaa 8100
Db 85693 ATTAGTGTCTTTCAGCCTTTGTATTGGGACAGTAGGCTCCATTAATTTTATGTGAAGTGA 85752
Qy 8101 aatlaggtatcggttcataatcagctctgtgattcttcacagcttccacatttcaactgt 8160
Db 85753 AATTAGGTATCGGGTCATTAATCAGTCTCTGATGCTTCCACAGCTTTCACATTTACCTGT 85812
Qy 8161 gataatcaagtgttcttctcaggtgttagccagagaaggggttcaggaactcttcc 8220
Db 85813 GATAATCAAGTGTGTCTTTCCTCAGGTGTAGCCAGAGAAGAGGGGTGAGGACCTCTCC 85872
Qy 8221 ctggtcgtagctttcactctgaaggttaaggaagacttccgagggctgtacatgactgg 8280
Db 85873 CTGGTCGTAGCTTTCATCTGAAGGTAAAGAGACTTTCCGGAGGGCTGTAGATGACTGG 85932
Qy 8281 ggtcctggtcagcactcctggttctgacttttcaatttaatttggaggttaggactcctg 8340
Db 85933 GGTCTGTGTGACGACCTCTGGTTTGCACTTTTTCATTAATTTGAGGTGAGCACTCTCTG 85992
Qy 8341 ttactgagacaagaagatagcagatcttccagaaaaagctgagtgaaggccgggtgcag 8400
Db 85993 TTACTGTAGACAAGAAGAGATAGCAGATCTTTCAGAAAAAGCTGATGGAAGCCCGGTGCAG 86052
Qy 8401 tggctcagcctgttaatcccagcacttgggagtcacaaagcaggtggatcacaggtcag 8460
Db 86053 TGGCTCAGCGCTGTAAATCCACGACTTTGGGAGTCCAAAGGCAAGGTGATCAGAGGTCTAG 86112
Qy 8461 gagttagaagacagcctgaccaaagctggtgaaacccctgctgtactaaaaatacaaaaat 8520
Db 86113 GAGTTGAGAACAGCCTGACCAACGTTGGTGAACCCCTGTCTGTACTATAAATAACAAAAAT 86172
Qy 8521 tagctgggtgtggtggcgcattcctgttaatcccagctacttgcagggcccaaggccaagaga 8580
Db 86173 TAGCTGGGTGTGGTGGCGCATGCCCTGTAAATCCACGCTACTTCCAGAGGCAAGCAAGAGA 86232

Qy	8591	atcgctgaacacacagggaagcgagggtgcagctgaggttgagattgcacaaattgcactccag	8640
Db	86233	ATCGCTTGAACACACAGGAGCGGAGGTTGCAGTGTGAGATTGCACCATTCGACATCCAG	86292
Qy	8641	cctgggtgacagagcaagacactctcaaaagaaaaaaaattcgcatagaagaatgcactgg	8700
Db	86293	CCTGGGTGACAGACGACACTCTCTCAAGAAAAAAAATTCGATGAAATGACACTGG	86352
Qy	8701	caatgagcctgcacaagaattactactgcactttcataattgtccatcacttgcagttt	8760
Db	86353	CAATGAGCCTGCACAAGATACTACTGACCTTTCATAATTGTCTATCCTGTAGGTTT	86412
Qy	8761	cagagtttagatgctctgtttctcaaaataaccocatacttttattctcttttaaat	8820
Db	86413	CAGAGTTTAGATGCTCTCTTTCTCAAAATAACCCATACTTTTATTTCTCTTTAAATTTT	86472
Qy	8821	tttccagtgccctgtccagcctccgtacaaattttttttttttttttttttttttttttt	8880
Db	86473	TTTCCAGTGCCTCTGCAGCCCTCCGTACATTTTTTTTTTTTTTTTTTTTTTTTGTGAGACCATGTCTGT	86532
Qy	8881	ctccatcgcttagcctggagtgtagtgcagtcggcacaaatctcgctcactgcagcctccacct	8940
Db	86533	CTCCATCGCCTTAGCCTGGAGTGTGCAGTGGCGCAAAATCTCGGCTCTACTGCACGCTCCACCT	86592
Qy	8941	cccaggttcaagtgatctctcgtcctcagcctcccaagtagctaggattatagtgcgcg	9000
Db	86593	CCCAGGTTCAAGTGATTCCTCGCTCAGCCTGCCAAGTAGCTAGGATTATAGTGTGCGCG	86652
Qy	9001	ccaccacaccagttaatttttgttattttttagtagagatggggtttcaacaatgttgcca	9060
Db	86653	CCACCACACCAGTTAAATTTTTGTATTTTTTAGTAGAGATGGGGTTTCCACCATGTGGCCA	86712
Qy	9061	ggcttgggttccactcctgcactcagtgatccaccacacttggcctcccaaatgctggga	9120
Db	86713	GGCTGGTTTTCACTCCTTGACCTCAGGTGATCCACCACCTTGGCCTCCCAAAATGCTGGGA	86772
Qy	9121	ttacaggcgtgaagcactgtgcctgggtccatatcttttatatttgcocatgattgtgtcc	9180
Db	86773	TTACAGGCGTGAAGCACGTGCGTGGTCCATATCTTTTTATATTTCGCAATGATTGGTCC	86832
Qy	9181	tttagaattcagaaaaattattgaaggcagctgtgtttgttttcttccatccatcaggc	9240
Db	86833	TTTTAGAATTTCAGAAATTAATTGAAGGCAGCTGTGTGTGTTTTTCCCTTCACATCCATCAGGC	86892
Qy	9241	cttatccaaagctctttaacctctgttttactcttatttcttccctccaatagctaaag	9300
Db	86893	CTTTATTCAAAGTCCTTTTAACTCTGTGTTTACTTTATTTTCATTTCCCTCGCAATGCTAAGG	86952
Qy	9301	tctaacaccagataattggaattatagctagcatccaaaggcctagatctgttaactc	9360
Db	86953	TCTAACACCAGATTAAATTGGAATATTAGCTAGCATTCACAAAGGCTTAGATCTGTAACTC	87012
Qy	9361	tgaatttggtccaaattccattaaaaatttttgttacaataagcgtgtttgttaagatctgac	9420
Db	87013	TGAATTTGGTCAAAATTCCATTAATAATTTTGTGTACAAATAAGCTGTTTGTAAAGATCTGAC	87072
Qy	9421	tagtggctatttttaataagaatttgcattaaaaattttatcaataaatttgaaacaaa	9480
Db	87073	TAGTGGCTTATTTTTTAATAGAAATTTTGCAATTAATAATTTTATCAATACAAATTTGCAACAAA	87132
Qy	9481	tttgtctaaaatgtgaaagatttccatgccttttttgggcttagattattttttaat	9540
Db	87133	TTTTGCTAAATATGTGAAAGATTTCAITGCTTTTTTGTGGGCTTAGATTATTTTTTAAT	87192
Qy	9541	gttgatttgaaatataatttggaaattgttatctaaaattctaaaagctacaaagtgaaata	9600
Db	87193	GTTCATTTTGAATATATTTTGAATTTGTAATTTGTATCTAAATTTCTAAAAAGCTACAAAGTGAAATA	87252
Qy	9601	ataatgaagtgaagttagttaattatagtggaagatcaattccagtagtatcttatctoga	9660
Db	87253	ATAATGAAAGTAAGTAGTGTAAATATTATGTAGTGGCAAGATCAATTGCCAGTATCAATTTCTATCGA	87312
Qy	9661	tttatpagaataatgtagattttcaLaaagtttaagtagtactgtttaaaggcgtattact	9720

	87313	Db	TTTTATTGAATAAGTGTGATTTTCATAAAGTTAACTACTGTTTAACAGCCTATTACT	87372
	9721	Qy	tgtatgcttcagtgatttagatagcaaaacattttttaagtgttttaaaaataatttatt	9780
	87373	Db	TGTATGTTTCGAGTWTAGATACAAANTCATTTTTTAAAGTTTAAAAATATTATT	87432
	9781	Qy	tttgataactatatatttatatttgtctgatttttaaacgttttctatcgtaaatctttaa	9840
	87433	Db	TTTGATAATCTATATTATATTGTCTGATTTTTAAACGTGTTTTCTATGGTAATCTTTAAA	87492
	9841	Qy	tcgtattcctgcgttcoggaataggtaacagtgagaatgatgaanaatgacaagctcaact	9900
	87493	Db	TCGTATTTCCTGCTTTTCGGATAGGTAAACAGTAGSAGCATGATGAAGAAGTGACAAGCTCACT	87552
	9901	Qy	tttacacactcgagcagdtgccctattatcaggcacgcgttcctggggctgcgcagctgc	9960
	87553	Db	TTTTACACACTCGGCAGTGGCCCTATTATACGGCAGCGCTTCCTGGGGGTGCCAGCTGC	87612
	9961	Qy	ctgcctgggctttcccatctcttccttgctgtctctcgcggctccctctgagggctgct	10020
	87613	Db	CTGCCCTGGCTTTTCCATCTCCCTTGCTGTCTTCTGGGCTCCTTCTGAGSGCTGCT	87672
	10021	Qy	gtcactggattagocctatabaocgtcttcocctctctaataatatttgtcgtctcaggtg	10080
	87673	Db	GTCACCTGGATTAGGCTATACGCCTTTCCCCTCTCTCTTAATTAATTGCTGCTCTCAGGTG	87732
	10081	Qy	aggttttgaaagcaataaagctagctaggtccaagttccagagctctctggcataggg	10140
	87733	Db	AGGTTTTGGAAGCAATAAAGCTAGGTAGGTCAAGTTCAGAGAGTCTCTGGCATGAGG	87792
	10141	Qy	acctgaaaaactcatctgtttggaagacctggctttggcagctggcgacctgtggggc	10200
	87793	Db	ACCTGAAAACTCATCTGTTGGAAGACTCGGCTTTGGCAGAGCTGGTGCACTGTTGGGC	87852
	10201	Qy	gttattggctgcgttctggctctcatcagttctccagataactctgcattctccagaaggg	10260
	87853	Db	GTTATTGGCTCGTTCGGCTCTCATAGCTWTTCCAGATACTCTGCATATCCTCASAGAGG	87912
	10261	Qy	aacatattccatgggttgagttcagctcccaggagatagggtttccctcccttaagtgc	10320
	87913	Db	AACATATCTCATGGGTGAGTTCAGCTCCCAGGGAGATGGGTTTCCTGCGTTAAAGTCG	87972
	10321	Qy	gcaagtaaccttttttctttttttgagacagagctctcgtctgcaccaggctggagt	10380
	87973	Db	GCAAGTACCCTTTTTTCTTTTTTTTGGACAGAGTCTCGCTCTCACCAAGGCTGGAT	88032
	10381	Qy	gcagtggtgcatttgtggctcaactgcacctctgcctcccagggtcgaagcagttctcctg	10440
	88033	Db	GCAGTGGTGCATCTTGCTCACTGCAACCTCTGCCCTCCAGGGTCAAGCAGTTCTCCTG	88092
	10441	Qy	cctcagctcccagtagctggagactcaggagcgaccacatgccagctaatttttg	10500
	88093	Db	CCTCAGCCTCCGAGTAGCTGGGACTACAGAGGCGCACACCATGCCAGCTAAATTTTG	88152
	10501	Qy	tatttttttagtagacgsggtttccacctgtttggccagagatggtctggaatctcttgat	10560
	88153	Db	TATTTTTTTAGTAGACGCGGGTTTCACCATTGTTGGCCAGGATGGCTGGATCTTTGAT	88212
	10561	Qy	ttcctgatcccgcttgccttggcctcccaaaagtgcgggattacaggcgctgagccatcatg	10620
	88213	Db	TTCTGTATCCGCCTGCCCTTGCGCTCCCAAAGTGTGGGATTACAGAGGCTGAGCCATCATG	88272
	10621	Qy	accagccttatgttattttgtttgttttcttgagatgagctcgcgtctgttgc	10680
	88273	Db	ACCAAGCTTTATGTTCTTGTGTTTGTGTTTCTTGAGATGGAGTCTCGCTCTGTTCG	88332
	10681	Qy	ccaggctggagtgagtggttgccatctcgacttactgcaacctgctcttccagggttcaa	10740
	88333	Db	CCAGCTGGAGTGCAGTGTGCCATCTCGACTTACTGCAACCTCTGCCCTTCAGAGTTCAA	88392
	10741	Qy	gtgattccttgccctcagcctcccgctgtagctgggattcacagggtgcctgcaccaatgccg	10800

Db 88393 GTGATTCCTTGCTCAGCCTCCCGTGTAGCTGGGATCATCAGGTGCTGCCACCATGCCG 88452
Qy 10801 gctaatttttttagtagacacaggttttcgccaatttggcgaagtagtctcga 10860
Db 88453 GCTAAATTTTGTATTTTAGTAGACAGGGTTTCGCCATGTGGCCAGGCTAGTCTCGA 88512
Qy 10861 actcctgacctcaagtgatcgtctcctcagcctcctaaagtctggtgggttacagaggt 10920
Db 88513 ACTCCTGACCTCAAGTGATCGCTTCCTCAGCCTCCTAAAGTGTGGGTACAGGAGT 88572
Qy 10921 gaacacacatgccagcctcaattacacctttattattattattattattttttg 10980
Db 88573 GAACACCATGCCAGCCTCAATTTACCTTTATTTATTTATTTATTTATTTATTTT 88632
Qy 10981 agacggagctcttctgttgcacaggtcgagtcgctgctgcgaactctagctcactgc 11040
Db 88633 AGACGGAGTCTTTCTGTGTGTTGCCAGGCTGGAGTGGGTGGCGCAATCTTAGCTCACATGC 88692
Qy 11041 aacctctctcctcaggtctcaagtgtattctcatgcatcagcttcccgagtagctgggact 11100
Db 88693 AACCTCCTCCTCCAGGCTCAAGTGATCTCATGCATCAGCTTCCCGAGTAGCTGGGACT 88752
Qy 11101 tcaaggctccgcacacacacttggctaatttttggtttttagtagagacggggtttcac 11160
Db 88753 TCAGGTGCCGCCACACACTTGGCTAAATTTTGTGTGTTTGTAGTAGACGGGTTTCAC 88812
Qy 11161 catgttgccaggtggtcttgaattctgacctcaaatgatcctcctgcttcagctccc 11220
Db 88813 CATGTGGCCAGGCTGTCCTGAAATCTGACCTCAATGATCTCCTGCTTCAGCCCTCC 88872
Qy 11221 caaagtcgtggattacaggtgagcactgcccacacagcaagtagtacccttttaaacat 11280
Db 88873 CAAAGTGTGGATTACAGGCTGAGCCACTGCCCCCAACAGCAAGTACCTTTTAAACAT 88932
Qy 11281 tagagacatttagtggccatcctcaaacccgtttgggtgtgtgagagaaagtgtggctg 11340
Db 88933 TAGAGACATTTAGTTGGCATCCTCAAAACCCGTTTGGTGTGTGGAGAAATGTTGGTGC 88992
Qy 11341 tgaatggtgttagtattcaagatgtcagcacaatcaatcaatcactgtgagatgcac 11400
Db 88993 TGACATGTTGTAGTTATCTAAGATGTACGCCATCAATCATCTGTGTGATGTGCAC 89052
Qy 11401 actgaagctgtaactctcatctagtagtatatttttaagatggaaatcttcaacccc 11460
Db 89053 ACTGAAGCTGTAACTCTCATCTAGATGATATTTTAAAGATGGAAATTTCTACAAACC 89112
Qy 11461 tgaataaagatttcagatccaaatttgagactcagccctacgagtaactctttaactt 11520
Db 89113 TCAGAAATGAAGATTTTCAGATCCAAATTTGAGACTCAGCCCTACGAGTAACCTTTAACTT 89172
Qy 11521 cagagagttaaaagaagatgcacagtgtgaagatttaagagaaatggaatcaaa 11580
Db 89173 CAGAGAGTTAAAGAAGATGACAGTGTATGAAGATTTAAAGGAGAAATGGAATCAAA 89232
Qy 11581 tgtcatttagcactcaaggcctacatgtcatttctgcacatttttctgttgtgtaaat 11640
Db 89233 TGTCTATTAGCACTCAAAGCCCTACATGTCTGTGACATTTTCTGTGTGTGAAT 89292
Qy 11641 ttttttttctataaaatgattgtgaagtgttctcgttagaattattgttgcctttcta 11700
Db 89293 TTTTGTTCCTATAAAATGATTGTGAAGTTTCTGTGTGAATTTATGTTGCCCTTCTA 89352
Qy 11701 atgtaataagcatttaggttttttttttttttttttttttttttttttttttttttttt 11760
Db 89353 ATGTAAATAGCATATTAGGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGAGACAGTCTCA 89412
Qy 11761 ctctcgtccaggtgtgagtgagtgagtgagcagatctcgggtcactgcgaatcttcgcctc 11820
Db 89413 CTCTGTGCCAGGCTGGAGTGCACTGGCAGCATCTCCGGTCACATGCAATCTTCCGCCTC 89472
Qy 11821 ctgggttctcgtcagcctcccgagtagtgggactagtcgggactcagcgacgtccacacccg 11880
Db 89473 CTGGGTTCTTGCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCACGTTCACACACCCG 89532

Qy 11881 gctaatttttttagtagtagacaggtattcacgctcttagccaggtgtgtcttg 11940
Db 89533 GCTAAATTTTGTATTTTAGTAGTAGACAGGATTCACCGTCTTAGCCAGGGTGTCTTG 89592
Qy 11941 atcctcgtacacctcatgatctacccgcctcgcctcccaaaagtgcgtggattacagggcatg 12000
Db 89593 ATCTCCTGACCTCATGATCTACCCGCCCTCGGCTTCCCAAAGTGTGGATTACAGGCATG 89652
Qy 12001 agccgctgtgcctggctattagagattttttattataatttatctccaaagataaaagcag 12060
Db 89653 AGCCGCTGTGCTGGCTATTAGAGATTTTATTATTAATTTATCTCCAAAGATAAAGCAG 89712
Qy 12061 tgaattattattggccacataattgaaataacaagagaaataaaatcatccatgctttt 12120
Db 89713 TGACATTAATTTGCCACATAATTGAAANAATACAAAGAGAAATAAANAATCATCCATGCTTT 89772
Qy 12121 gttagcctcactgtcattgaaattattgttacatggcagtttgcgtgtgtgtctc 12180
Db 89773 GTTAGCCCTATCAGTCTGATTTGAATATTATGTTACATGCGAGTTTGTCTGTGTGCTC 89832
Qy 12181 tgttaggcaagcctcgtgtgacattcctttagctatttaattgaggaattagaaatgacag 12240
Db 89833 TGTTAGGCAACGCTCTGTGACATTTCTTTAGCTATTAAATTGAGGAATGTAGAATGACAG 89892
Qy 12241 aacagtgcttctcctcaatgatacttgaggaattattttaacttccatattgaaattaca 12300
Db 89893 AACAGTGTCTCCTCAATGATCTTGAAGGATATTATTAACTTTTCATATTGAATTACA 89952
Qy 12301 ttttataaaattataatagtttaagtctgggaataaaacactgatttaagtcatttg 12360
Db 89953 TTTTATTAATTTATAATGAGTTAATGCTGGGAAATAAACAACACTGATTTAAGTCATTTTG 90012
Qy 12361 gcttttagtactaaagcatttgacaataaaatgaacttcttcagaataataggatacactctg 12420
Db 90013 GCTTTTAGTACTAAAGCATTTGACAATAAATGACTTCTTCAGAATATGATATACCTCTCG 90072
Qy 12421 aaagcaataaacgcattttaatgaattgtlaaggaataaacatcattttatttttatttt 12480
Db 90073 AAAGCAATAAACGCATTTTAATGAATTTGAAGAAACAACAATCATTTATTTTATTTT 90132
Qy 12481 ttttttgagacagacttctgcttttggcttaggtggagtcgaatggcgagctctg 12540
Db 90133 TTTTGTTCAGACAGACTTTTCGCTTTTGTTCCTTAGCTGGAGTGAATGGCGCATCTCG 90192
Qy 12541 gctcactgcaacatccgctcctggttcaagagattctcctcctcagctcctcaggttg 12600
Db 90193 GCTCACTGCAACATCCGCTCTGTGGTTTCAAGCGATTCTCTCCCTCAGCTTCTCTGAGTTG 90252
Qy 12601 ctgggattacagggcagtcgccaccacgcctggccaattttgtatttttagtagagacgg 12660
Db 90253 CTGGGATTTACAGGCACGTGCCACACGCTGGCCCAATTTTGTATTTTGTAGTAGACAGCG 90312
Qy 12661 gttctccacgtgtgtcaggtgtgtctcaaacctcctgacctcagtgatctgtccgcctc 12720
Db 90313 GTTTCCTCACGTTGTGTCAGGCTGGTCTCAAACTCTCTGACTCAGGTGATCTGCCGCCCTC 90372
Qy 12721 agctcccaagtagtactggaattacagggcgtgagccacgcgtgctcgtggcacaataattatt 12780
Db 90373 AGCTCCCAAGTAGTCTGGAATTTACAGGCGTGAGCCACGCTGCTGCCCAACATTTATTAT 90432
Qy 12781 ttttttttaattctagaaaaatacaccttctaaagaaatttgattaaaaaccaacctcttctcatt 12840
Db 90433 TTTTATTAATCTAGAAAAAATACACTCTTAAGAAAAATGATTAATAAACCAACCTTCTTCATT 90492
Qy 12841 agccctcaagatcaacatctattgttctcttgttgcagcttccacttaagagagagaaactt 12900
Db 90493 AGCCCTTAAGATCACATCTATGTTCTCTTTGTTGAGCTTCACTTAAAGAGAGAAACTT 90552
Qy 12901 ctgtgtctcaagcaagagagatggcctcccgatgtgagtggtggggtaagcctct 12960
Db 90553 CTGTGTGCTCAAAAGCAAGAGATGGGCTCCCGAGTGTGAGTGTGGGGGTAAAGCTTCT 90612

QY 12961 ggggactcactgggtacacactgtccacttaaggaatacatttcacagagcgcttgccct 13020
|||||
Db 90613 GGGGACTCCTACTGGGTACACTGCTCCACTTAAGGAATACATTTACAGAGGCTTGCCCT 90672
QY 13021 ctttcatttcagtgggacagctgccatcgctcccatcattgctgctgcacgaagcgagaa 13080
|||||
Db 90673 CTTCAATTTGAGTGGGACAGCTGCCATCGCTGCCATCATTTGCTGCTCAAGGACGGGAAA 90732
QY 13081 aqcatcaactcatgaaggaagaggtgaatgcctcggttttcttgatncagcagttacag 13140
|||||
Db 90733 AGCATCACATCATGAAGGAAGAGAGGTGAGATGCTGGTTTCTTGATTCAGCAGTTACAG 90792
QY 13141 gtagggtctgaatgctgggcagagctgctctcttcacggccctacagacacacactttg 13200
|||||
Db 90793 GTAGGGTCTGAAATGCTGGGACAGTCTGCTCTTCAGGCCCTACAGACACCACCTTTG 90852
QY 13201 aaggacgttggaacagtttggaatcacactcagctaaagtatgataaaatggccctcttttatctg 13260
|||||
Db 90853 AAGGACGTGGAAACAGTTTGGACATCACTCAGCTAAGTGATAAAATGGCCCTCTTTATCTG 90912
QY 13261 tgtttgtcccgcatgtcaacacggctgcaattcgagcattttgtagattgtccatttag 13320
|||||
Db 90913 TGTTTGTCGGCATGTCAACACGGCTGCATTCGAGCATTTTGTAGATTTGCCATTTAGG 90972
QY 13321 atctagtcaccgtccctcttaagggtgcatgctcttccttggttaacttgagctcagagacag 13380
|||||
Db 90973 ATCTAGTCACCGTCCCTTAAAGGTGCATGCTTTCTTGGTACTTGAGCTCAGGACAG 91032
QY 13381 tgtctaaacacagaccoccatatggatgggcctgggggttttatgggtccagaggaatgccaca 13440
|||||
Db 91033 TGTCTAACAACACAGACCCCATATGGATGGGCCCTGGGGTTTATGCTCCAGAGGAATGCCACA 91092
QY 13441 gtattctatgtcaaatatttctctgactctgagacattagaccagtgaccacaga 13500
|||||
Db 91093 GTATTCTATGTCAAGATATTCTCTGACTCTTGAGGACATTAGGACCACTGAGTGGCCACAGA 91152
QY 13501 ctgaagaaacaccttaatgccaagctcccttctggccagtgtaggcctgaagtgccctca 13560
|||||
Db 91153 CTGAAGAAACCTTAATGCAAGCTCCCTTTCTTGCCAGTGTAGGCCCTGAAGTGCCCTCA 91212
QY 13561 acctgacagttaacctgtttaggtatccacaagagacacgaaggggtgtgtgatgtg 13620
|||||
Db 91213 ACCTGACATTAACCTGTTTATAGTATATCCAAAGACACAGAGAGGGGTGTGATGGTGATGT 91272
QY 13621 gtaaaagtgtgtgtgtgtgttttaccctcagctcactcgatagatatgtcatgttca 13680
|||||
Db 91273 GTAAAGTGTGTTTTGTGCTTTGTTTACCTCTCAGCTCACTGGATGGATATGTCATGTGTA 91332
QY 13681 gcagttgccttgaagcgagttcagtttggctggctgagctgtgacccccagtgggcggtct 13740
|||||
Db 91333 GCAGTTGCCCTTGAAGGCAGTTCAGTTTGGTGGCTGAGCTGTGACCCCAAGTGGCGGGCT 91392
QY 13741 tatttggttttcagatatttgctgaagagctgtgtactctccagatcctggtgctgt 13800
|||||
Db 91393 TATTTGGTTTTGCAGATTTTGGCTGAAGAGCTGTGTACTCTCCACAGATCCTGGTGTGCT 91452
QY 13801 ttgttggtgtagaatgtccagataaaagcttcatccaaccatctgtgagaatgccacc 13860
|||||
Db 91453 TTTGTGGTGGTGAATGTCCAGATGAAGCTTCAATTCACCCATCTGTGAGAAATGCCACC 91512
QY 13861 tttcagaggttaatgaggggtctctcaggttgggagaagtgagagctgaacccagccagc 13920
|||||
Db 91513 TTTTCAGAGCTAATGAGGGGTCTCTAGGGTGGGAGAGTCAGAGCTGAAACCCAGCCCCAGC 91572
QY 13921 atcgacatagggcatcttgtgcaagagctgtgtttctgggaagaccactatctgggttca 13980
|||||
Db 91573 ATCGACATCGGCATCTGTGTGGCAAGAGCTGTGTTCTGGGAAAGACCACTATCTGGGTTTA 91632
QY 13981 cagttcagagggcggaacctcctgccttaagctcactgttagttggtgggtcccggtgt 14040
|||||
Db 91633 CAGTTTCAGAGGGCGGCACCTCCCTTAAAGTCACCTGTTAGTGGTGGGTCCGGTGT 91692
QY 14041 acacagctcctcaaytgaattagaaaaagattgaaaaactgaacaaactgaggaactagaaa 14100
|||||

Db 91693 ACACAGCCTCAAAGTGAAATAGAAAAGATTGAAAACACTAGAAAACACTGAGGACTAGAAA 91752
QY 14101 ttcaactagaactcttaacagctcttataccagaagaatacttagaacttttttgaattct 14160
|||||
Db 91753 TTCAACTAGAACCTCTTTACAGCTCTTATACAGAAGAAATCTAGAACTTTTGAATTTCT 91812
QY 14161 aactaatgcccagattatcatttgattattttgaactgaataaattttcttccattac 14220
|||||
Db 91813 ACTTAATGCCCAGATATCATTTGGATTATTTTGAACITGAATTAATTTCTTCCATTTAC 91872
QY 14221 ctgcattgaaacaaaatgagtggtgcagagtggtgagagactgctgggtcaagagtcctgt 14280
|||||
Db 91873 CTGCATTTGAAACAAATAGAGTGGGTGCAGAGTGTGCAGACTGTGCTGGTCAAGAGTCGT 91932
QY 14281 gttatgggtatggaactcaacagctggggaatgtcttttgggtaactgccactctgtgttg 14340
|||||
Db 91933 GTTATGGGATGGACTCACAGCTGGGGAATGCTTTTGGGCTAACTGCCACTCTGTGTGTG 91992
QY 14341 tctctctatcgaagttaaccagttttgcggttcagcttccattccagatggaatcatttt 14400
|||||
Db 91993 TCTCTATCGAAGTTAACCAGTTTTGCGGTTCAGCTTTTCATTCAGATGGAATCACTTTT 92052
QY 14401 gacccacctatctgagtttgaatcttttcccccaactcttaatgggtttacactgtattttc 14460
|||||
Db 92053 GACCCACCTATCTGAGTTTGAATCTTTTCCCCACCTCTTAATGGTTTACCTGTATTTTTC 92112
QY 14461 ctgttctcagtttgatctatctgtatttttcaactgttttttcttaacttaacacaaca 14520
|||||
Db 92113 CTGTTCTAGTTGTATCTATCTGTATTTTTCATCTGTGTTTTTCTACTTTTACCAACA 92172
QY 14521 aatccttttggcgctgtaccccttccagtcagagctcagagcttagagagttgtttcatggtct 14580
|||||
Db 92173 AATCCTTTTGGGCTGCTGTACCCCTTCCGAGTCAGAGCTTAGGAGTTGTTTCATGCTCT 92232
QY 14581 gcttattctctgtgggtgaatttggatcgctggtagccccggcttgttattttaatcc 14640
|||||
Db 92233 GCTTTTATCTCTGTGGGTGAATTTGGATGCGTGTGTGAGCCCCGGCTTTGTATTTAATCC 92292
QY 14641 agttttgggcagcaaaacacttccaatgaatcaggtgtctatttgagcagcatgtggat 14700
|||||
Db 92293 AGTTTTGGGCAGCAAAACCTCTTCAATGAATCAGGTGTCAATTTGAGAGCCATGTGTGGAT 92352
QY 14701 gtgtgatgtgtgtggatagataaaaaatgactactgtatatttctttttaaagggaac 14760
|||||
Db 92353 GTGTGATGATGTGGGATAGATAAANAATAGCTACTGTGTATATTTCTTTTAAAGGGAAC 92412
QY 14761 tggagggaacacacatcagcatgttagtaagtggtctgtttgtccagtggtgaaatttcag 14820
|||||
Db 92413 TGGAGGGAACACATCAGCATGTTAGTAAGTGGTCTGTGTCCAGGTGGTGAAATTTTCAG 92472
QY 14821 atgattttcatittctgtctgtctcaggtcctctctggaagagacacacaggtggc 14880
|||||
Db 92473 ATGATTTTTCATTTCTGCTGCCCTGTGCTCAGGTCTCTGGAAGGCAGACACAGGTGGC 92532
QY 14881 attgaggtgcaggaagtttatttcgaggaatttctgactgtgagagaggaagagaggg 14940
|||||
Db 92533 ATTGGAGGTGCAGGAGGTTTATTCGAGGAAATTTGACTGTGAGAGGAGGAGAGAGGG 92592
QY 14941 agcagaggtgcagggagagcctgggtctggcttgcaggttggaccctgtatgagtgga 15000
|||||
Db 92593 AGCAGGAGGAGGAGGAGAGGCTGGGTCTGGCTTTTGCAGTTTGGACCCGTATGAGTGGA 92652
QY 15001 gagggttagaagaagaagtgagtgctgagaaagatcagcaggcctacttggaagccag 15060
|||||
Db 92653 GAGGGTAGGAAGGAAGTGCAGTGTGAGAAAGGATCAGCCAGGCCCTACTTGGAAAGCCCCAG 92712
QY 15061 agcagagcttgcagatacaggaatccccgcgtccatttgaaatggccccagcacccgggtc 15120
|||||
Db 92713 AGCAGAGCTTGCAGATACAGGAATCCACGTCCTCATTTGGAATGGCCAGCACCGGGTC 92772
QY 15121 tgcctgtagcagcctgtctgtgagagcatggcctgggctgtggaggtctgcagctcactgca 15180
|||||

Qy	-	17341	ttttattattattattatttttttttgagacggagtttcaactctgtgtgcccaaggcgtgaggt	17400
Db		94993	TTTTATTATTATTATTATTATTTTTTTTGTGAGACGGAGTTTCACTCTCTGTGCCACAGGCTGGAGT	95052
Qy		17401	gcaatggcgtgatctcaqctcactgcgaactccgcctcccggttccaagattctcctg	17460
Db		95053	GCAATGGCGTGATCTCAGCTCACGTGCACAACTCCGCCCTCCCGGGTTCAAGCAATTTCTCCTG	95112
Qy		17461	cccagcctcccgaaatagctggaattacaggcgctgtaccacgcgcggcgtactattttt	17520
Db		95113	CCCCAGCCTCCGAATAGCTGGAATACAGSGCCTGCTACCAACGCCCGCTAATTTTTT	95172
Qy		17521	gtaatttagcagagacgggtttcaccaatgttgacggctgtgcagacccctcgaacttc	17580
Db		95173	GTAATTTTGTAGCAGAGACGGGTTTCAACATGTTGCGAGGCTGGTTCGAACCTCTGACCTC	95232
Qy		17581	aggtgatccaccaccctcagcctcccaagtgttggaattacaagcgtgacgtgccg	17640
Db		95233	AGGTGATCCACCCACCTCAGCCTCCCAAAGTGTGGATTTACAGGGCTGAGCGGTGCGCG	95292
Qy		17641	gcctgttttaatttttaaggatctgaacctgtatttaagtttctcctcccaactccacagt	17700
Db		95293	GCCTGTTTAAATTTTAAAGATCTGAACCTGTGATTTAAAGTTTCTGCCACCTCCACAGT	95352
Qy		17701	atttgtattagaatagagcatgtgctggattatgactggatgtgtgtgtgttgaggtt	17760
Db		95353	ATTTGTATTAGATAGACATGTCTGGATTTATGACTGGATGCTGTGTGTTGTGAGGTT	95412
Qy		17761	gggtagttggggccctttaagagactatactagcaagactcgggccacaggaacaatca	17820
Db		95413	GGGTAGTTGGGGCCCTTTAAGAGACTATCTAGCAAGACTCGGGCCACAGGCAACATCA	95472
Qy		17821	cggggttgaagaacctgggtgtccctttgttggcatctgcgcaggtctctaacacacagca	17880
Db		95473	CGGGTGTGAAGAACCTGGTGTCCCTTTGTTGGCATCTGCGCAGGCTCTTACACACAGCA	95532
Qy		17881	gcgatacacagccctagccacattcagattaccttgttgotgtgaaaaattgtgcaca	17940
Db		95533	CGGATACACAGCCCTAGCCGACATTCAGATTACCTTGTGCTTGTGAAAAATATTGCACA	95592
Qy		17941	gggcctgccctagacctagtaattagaactcttgagagttaggcttgggactcaaaagct	18000
Db		95593	GGGCGTCCCTTAGACCTTAGTGAATTTAGAAATCTTGAGAGTTAGGCTTCGGACTCAAAAGT	95652
Qy		18001	cccagatgattttaatgctcagcgaggttgaagagcgccctgtccaagaggtgtccactc	18060
Db		95653	CCCAGATGATTTTAAATGCTCAGCCAGGTTGNAGAGCCGCTGTCCAGGAGTTGCCACTC	95712
Qy		18061	cgtgtatctggggcttcttaggaaatgggactcagggcctcaactcagagctgccgaa	18120
Db		95713	CGTGTGATCTGGGGCTTGCTAGGAAAGTGGGATCTCAGGCCTCACTCAGAGCTGCCGAA	95772
Qy		18121	ctggctctgcgttttgccaaagttctgggtgtaacatgagtttcagagtaactcctc	18180
Db		95773	CTGGCTTCTGGGTTTGTCCAAAGTTCTGGGTGTGAACATGAGTTTCAGAGTCACTCCTC	95832
Qy		18181	tagggcccttctctcagctcgaaccattgacccctcaagacacatttgcaaacatctg	18240
Db		95833	TAGGGCCCTCTCTCTCAGCTCGGACCATTGACCCCTCAGAGGACATTTGGCAACATCTG	95892
Qy		18241	gaaacgttcttggttcacagcctagaggtgggtagtgtgctgctagtgggtagaggtg	18300
Db		95893	GAAACGTTCTTGGTTGTACACGCCTAGGAGGTGGGTAGTGGTGTGCTAGTGGGTAGAGG	95952
Qy		18301	tcaggggtactgcacacaggacagcagcactggccacagaaaaaactgtcttgccctgag	18360
Db		95953	TCAGGGGTACTGTCACAGGACAGCAGACTGGCCACAGAAAAAATCTCTTGTGCCCTGAG	96012
Qy		18361	catcagtagttcccggttgactgacctgagcagacagcattgacatccaaaagcggt	18420
Db		96013	CATCAGTAGTTTCCCGTTGACTTGCCCTGAGGACAGCGATGCAAGATTCACAAAAGCGGT	96072
Qy		18421	ggagcagacctgcbccagatcctagtcaacttaaccttcagttgttgactgaagaaacttc	18480

[illegible]

Db 97153 CCCCCTCGGAAGACAGACATTTCTTACCTGGTAGGGCTGGGAGGATCGGACCCGAAGTGCA 97212
QY 19561 tggcttgagtgagagctggatgcaaggcacacactgttctcttgaataaaatgacagt 19620
Db 97213 TGGTCTTGCAAGTGAAGCTGGATGCAAGGACACACACTGTTCTCTTGAAATAAATGACAGT 97272
QY 19621 tcccagcataaagaataatgtcatcttttttaaatgtaaaaagaattacagcaaatctctttgaag 19680
Db 97273 TCCGAGCATAAAGAAATGTCAATTTTAAATGTAAAAGAAATTACAGCAATTCCTTTTGAAG 97332
QY 19681 aaaggactggagaatttatattgtctctcttagccttttggtagacagatagcctgtggctc 19740
Db 97333 AAAGGACTGAGAAATTAATTTGTTCTTGAAGCTTTTGGTACAGATAGCCTGTGGGTC 97392
QY 19741 ccacactggcgaaagcccttggtttcagagcaggttcaccaggggccctgccagtcctccctc 19800
Db 97393 CCACACTGGTGGGAAGTCCCTTGGTTTCAGAGCGGTGCCAGGGCCCTGCCAGTCCCCCTC 97452
QY 19801 ctgggaagctggatagaaactatgttgcttaccctctgtcttagctgtgttttggttatt 19860
Db 97453 CTGGGAAGCTGGATAGAACTATGTTGCTTACCCCATCTGCTTAGCTGCTGTTTGTATT 97512
QY 19861 ataaaaaataatgtgagactgaggaatattatcaagaaaaaagagtgatattagtttcacgg 19920
Db 97513 ATAAAGAAATATGTGAGACTGGGTAAATTTATCAAGAAAAGAGGTGTATTTAGTTTACGGT 97572
QY 19921 tctgcggctgagaattgaaggtcacggccctagcttccagtagaagcttccatgcctgca 19980
Db 97573 TCTCGGGCTGAGAAATTTGAAGGTCAAGGCTCAGGGCCCTAGCTTCCAGTGAAGGCTTCCATGCTGCA 97632
QY 19981 tcaatacgtggcagagagcgcaagtaggaagtgagcgcttggtaagcgggataaacctg 20040
Db 97633 TCATAACGTGCGAGAGGAGCGCAAGTAGGAAGTGGACGCTGTGAAGCGGGATAACCTG 97692
QY 20041 agctgactctggtttataaacaacccccctctcctctgggaacaaatccatcccttgaga 20100
Db 97693 AGCTGCACTCTGGCTTTATTAACAACCCCTCTCTCTGGGAACAAATCCATTTCCCTTGGA 97752
QY 20101 agtaatcagctctctgagacagacttactactactagctcacaagccactcaggagg 20160
Db 97753 AGTAATCAGTCTCCTGAGAGCCAGTACTTACTACTGAGCTCCAAGCCACTCAGGAGG 97812
QY 20161 tccgtccctgtagcccaaacgcttccactaggcccgctcccaaacgcccatactag 20220
Db 97813 TCCGTCCCTGTAGCCCAACGCTTCCACTAGGCCCGCTCCCAAAACCCGCTACTAG 97872
QY 20221 ggagcagtttccacatgaggtctggggacaacaaatgacactcaaaccaattgcactt 20280
Db 97873 GGAGCAGGTTTCCACATGAGGTCTGGGGACAACCAATGACACTCAAAACCAATTGCACCTT 97932
QY 20281 ctcatggctgcatggtgctcaacttttgaccccaaggaatgattgtttcacatggattt 20340
Db 97933 CTCATGCTGCATGCTGGCTCACTTTTGACCCAAAGAAATGGATTTGTTTCATGGAATTT 97992
QY 20341 ttccagagaaaaagagtcagtagccagaaatcatcttcttggaaaggtctgccaat 20400
Db 97993 TTTTCACAGAGAAAGAGTCACTACCCAGAAATCATCTTCTTGGAAACAGGGCTGCGCAT 98052
QY 20401 cccgatgaattcgaattcagtagcagcaacttgcacaaatggaatgctgtttccc 20460
Db 98053 CCCGATGAAGATTCGAAATGTCACTGCGCACACTTGTCAACATAAGGTATGCTGCTTTCCC 98112
QY 20461 aggaagcatccttccatcaagggcaygtttactttttaacaaagtcctgctgtaactca 20520
Db 98113 AGGAAGCATCCTTCCATCAAGGGCAGCTTTACTTTTAAACAAAAGTCCCTGCTGACTCA 98172
QY 20521 ccagtcatttgaatcggtatcaagccctgtcacctgtcatctgactgagtgaggtcca 20580
Db 98173 CCAGTCGATTGGAATGCGGTATCAAGCCCTGTCACTTGTCTATGCTGACTGGAGTGCCA 98232
QY 20581 ggagaggagcgtggccttactgattttatagcctcagtagcaaaccttaccctgggaat 20640
Db 98233 GGAGGAGCGTGCCCTTACTGCAATTTATAGCCTCAGTAGCAAACTTACCCTGGGAT 98292

QY 20641 caccaaaattccatcccatgatgtcttttaataaacagactgatttttactgtgggcagtaca 20700
Db 98293 CACCAAAATTCATCCCATGATGCTTTTAATAAACAGCTGATTTTACTGTGGCAGTACA 98352
QY 20701 cctagctaagaaattagctcctttaatttttacattaatcctatgaagcgtggaataact 20760
Db 98353 CCTAGCTAAGAAATTAGCTCCTTTAAATTTTACATTAATCCTATGAAGTGGTGAATAACT 98412
QY 20761 acccatittgtgatgagtgacctgatatccagagagtgactgctgattgctctacag 20820
Db 98413 ACCCATTTGTTGATGAGTGACCTGATATCAGAGAGTGACTTGCTATGGTTCTCTACAG 98472
QY 20821 ctggtaaagtgggcatctgaaagtttgagcgggacttgggtcttcttgatgctacacgtga 20880
Db 98473 CTGTTAAGTGGGGCATCTGAAGTTTGACGGGACTTGGGGCTTGTGATTTGCTACATGTA 98532
QY 20881 ttgtccccagccatttggtagtagttgtaaaagctttaggggttttggacatttgg 20940
Db 98533 TTGTCCCCAGCCCATTTGTGGTAGTAGTTTAAAGAGCTTTAGGGTTTGGACATTTGTG 98592
QY 20941 ttcagaaccttattggattcccttgacatgatttttagttgattctctctgggtttgccc 21000
Db 98593 TTCAAGAACCTTTATTTGGATTCCTCTGACATGTTTTTAGTTGATTTCTTTGGGTTGCC 98652
QY 21001 tggggtcatcagcagagagattagtcacaaatcggttgacatgtacacggttatctctaca 21060
Db 98653 TGGGGTCTACACAGAGAGATTAGTCAAAATCGGTGTGCACATGTACAGTTATCTCTACA 98712
QY 21061 gatagtagtaagaaataagattgtgaattaccaggtttgttttaatttttgcctgc 21120
Db 98713 GATAGTAGTCAAGAAATAAGATTGTGAATTTACCAGGTTTGTTTTAAATTTTGTCTGTC 98772
QY 21121 catottacatctagtggtgatgataaaacaaacaaatagtgcatataataatatacagca 21180
Db 98773 CATCTTACATCTAGTGGTGATGATAAACCAACCAATAGTGCAATTAATATATACAGCA 98832
QY 21181 gtgcagagatgctcctgacatcagaaataacaaatcggttggtgttctctcttggagt 21240
Db 98833 GTGACGAGATGTCCCTGACATCAGAAATATACAATCTGGGGTGTGTTCTCTGTGGATG 98892
QY 21241 aggcacatgaataagcagcttgagtgagcggcctcccgggggtgagatcctggg 21300
Db 98893 AGGACATGCAATAAAGCAGCTTGGAGTGAGCGGGCTCTCCCGGGGCTGAGATCCTGGG 98952
QY 21301 ggaagagggtttttttagattgacctgacacctgagagagcttttgaacacagctgaa 21360
Db 98953 GGAAGAGGGGCTTTTGTAGTTTGACCTGACACCTCGCAGCAGCTTTTGAACACAGCTGAA 99012
QY 21361 gctaattgggaaggtgctattgccaaccttgcctccgctccgactcctttttcccccaaga 21420
Db 99013 GCTAATGGGAAGGTGCTAATGCCCTTGGCTTCCGCTTCCGCTTCCCTTTCCTTCCGCCAGA 99072
QY 21421 aggtlaattgcttagcaccggggtctctctgcaaaaatgggtgagcctctcagtgcttc 21480
Db 99073 AGGTAAATGCTTAGCACCGGGCTTCTCTCTGCAAAATGGGTGACAGCCTCTCAGTGTC 99132
QY 21481 gtggtcctccacagagaatgaaggaggccagaggggtcagcactctctctgcttggag 21540
Db 99133 GTGGCTCTCCACAGAAATGAAGGAGGCGCAGAGCGGTGAGCAGCTCTCTCTGCTTGGAG 99192
QY 21541 cagagcttctgaatgagctcacagagaatagcccaagaagtgtgtcagaatccagac 21600
Db 99193 CAGAGCTTCTGAAATGGGACTCACAGCAGAAATAGCCCAAGAAAGTTTGCAGAAATCCAGAC 99252
QY 21601 ttccagagccctgcttaaaacaaagtcagaaaccccgagtgacacctggaggtctgctt 21660
Db 99253 TTCCAGAGCCCTGCTTAAACCAAGTCAAGAACCCCGAGTGACACCTGGGAGTCTGCTT 99312
QY 21661 aactggctccctgaatgaagcaccttgagcccgccctgcaccaggtgtcttcttggagacat 21720
Db 99313 AACTGGCTCCCTGAATGAAGCACCTTGACGCCCGCCCTGCACACCGGTGCTTTTGGAGACAT 99372

Db 101533 GTGATTGGAAAGAGTAAAGGGGCACAGCCCGAGGCGATCATGGGGCGAGGTGGGGAGAG 101592
Qy 23941 agctgagagccctccagccaccccttccagtttccagacctgtctggtgcggcaactgca 24000
Db 101593 AGCTGACAGCCCTCCACGCCACCCTTTCAGTTTCAGACCTGCTGTGTGGCGGCACTGCA 101652
Qy 24001 agcatgcttggctgtgctgtgtgcacacctctggctggaagtgtctatttcgggg 24060
Db 101653 AGCATGCGTTTGGGTGTGGCTGTGTGCACAGCTCTGCGCTGGAAAGTGCTATTCCGGGG 101712
Qy 24061 acacatccctgcagagctgtgtccgagtgggtgagttagaggaagcaagccaccc 24120
Db 101713 ACACCATGCCCTGCAGGCTCTGTGTCCGGATGGGTGAGTAGAGGAAGCAAGCCACCC 101772
Qy 24121 tgaggttctctgggttgtgtagctgaggtgaatgcaggtggcttcagggaaacg 24180
Db 101773 TGAGGTTGCTCTGGGTTTGTGTAGCTGGAGGTGAATCAGGTGGCTTGCAGGGAACG 101832
Qy 24181 tcagcagaggcagagactcagggtcccccacctcagagttctctgtgtctatctcctagtag 24240
Db 101833 TCAGCAGAGGAGAGACTCAGGTCCCAACCTCAGAGTCTGTGGTTGTCATCCTAGTAG 101892
Qy 24241 gcaacccagggcagggagactgagttgttagaccaggaacacagcagtgactgaggcc 24300
Db 101893 GCAGACCCAGGGCCAGGGGAGCTGAGTGTGTGAGACAGGAACACACGTGACTGAGGCC 101952
Qy 24301 tgtgtgcgctctgcagagaactctgccctgactctgtgctgtctctccagggaaaga 24360
Db 101953 TGTGTGCCGCTCTCGCAGAGAACTCTGCCCTGATCTTGTGCTGTCTCCAGGGAAGA 102012
Qy 24361 tgcacccctctgatacatgaagccacctggaagatggtttggaaggaggcagtgga 24420
Db 102013 TGCCACCCCTCTGATACATGAAGCCACCTCGAAGATGGTTGGAAAGGAAGCAGTGGA 102072
Qy 24421 aaagacacaggttagcaaaagccgctcagtcctgtgtgcacacatcctctccctcccc 24480
Db 102073 AAAGACACAGGTAGCAAAAGCCGGTCAGTCTTGTGCGCCACATCTCTCCCTCCGCC 102132
Qy 24481 actactgacactgagcagcgtgtgtgtctccactgatgtggggctgcctgtctct 24540
Db 102133 ACTAGTCACACTGAGCAGCCGTGTGTGTCTCCACTGATGTGGGGCTGCCCCTGCTTCT 102192
Qy 24541 atcaagggtatgggggtctcttgacctgtggcagtgctcacaggctctctggcctttat 24600
Db 102193 ATCAAGGGCTATGGGGCTTCTCTTACCTGTGGCAGTCTCACAGGCTCTTGGCCCTTAT 102252
Qy 24601 ttttgcagaattttctaagcaagattctagagtgaaggcacagttttttgaaagcatctag 24660
Db 102253 TTTTGCAGAAATTTCTAAGCAAGATTTCTAGAGTGAGGCACACAGTTTTTTTGAAGCATCTAG 102312
Qy 24661 aaatcggtgaataaactataagccatgtcaggaattgccagggaaagcgagggtctgg 24720
Db 102313 AAATCGGCTGAATAAACHATTAAGCCATGTACAGGAAATTTGCCAGGGGAAGCGGGGCTGG 102372
Qy 24721 gggactgaatttttggctgctaatttcacgaaagatgcatcaccaggtggggccttg 24780
Db 102373 GGGACTGAATTTTGGCTGCTAATTTCAACAAAGAGTGCATTACCCACAGGTGGGCCCTG 102432
Qy 24781 tggttctcttgggtgcctcatggaacagatttggcagccagcagaggggtgggcttca 24840
Db 102433 TGGTTTCTCTTGGGTGGCCCTCATGGACAGATTTGGCAGCCACACAGAGGGTGGCTTCA 102492
Qy 24841 tccaggggtgtgtcgaaggctctggccctcaggggagattgtgctgctacgaggtgc 24900
Db 102493 TCCAGGGGTGTGTGGAAGGCTCTTGGCCCTCAGGGGAGATTGTGCTGCTACGGAGGTGC 102552
Qy 24901 ccgttaagaaacccaccagcttcccgggtgcccctggcagttgatggccaggggtctgtg 24960
Db 102553 CCGTTAAGAAACCCACACAGCTTCCCCGGGTGCCCTGGCAGTTGATGGCCAGGGTCTGTG 102612
Qy 24961 ccactgtctgttgcagttcttgcagttgagttcaggttcaggtctcaggtctcctcacc 25020
Db 102613 CCAGTGTCTGCTTTGCAGTCTTGCAGTTGAGTTTCAAGTTCAGTCTCAGTCTGCTCTGCTTCACC 102672

Qy 25021 tgcagcaaaogtccccaaagccatcagcgtggggtatgcgatatgaacgcgaggttcattatg 25080
Db 102673 TGCACACACAGTCCCAAGCCATCAGCGTGGGGATCGGATGAACGCGAGTTCATTATG 102732
Qy 25081 ctgaaccacttcagccagcgtatgccaaaggttcccccttcttcagcccccaacttcagcgag 25140
Db 102733 CTGAACCACTTTCAGCAGCGCTATGCCAAGGTCCCTCTTCAGCCCCAACCTTCAGCGAG 102792
Qy 25141 aaagtggaggttgcctttgaccacatgaaggtctatgtatgtcacagcagacacagggcg 25200
Db 102793 AAAGTGGGAGTTGCCCTTTCACCAATGAAGGTCTGTATGTACAGGACACAGCGGGC 102852
Qy 25201 gggacggggcagggagacaggtctcacactgagtagcaggtcagctgagtttgc 25260
Db 102853 GGGACGGGCGAGGAGACAGGACTCTACACACTGAGTAGGACGGTCAGCTGGAGTTGCT 102912
Qy 25261 ttcttatttggggccaccgttgggaaaggttatctaccatcactaacccaggtcgaaca 25320
Db 102913 TTCTTATTTGGGCCACCGTGGGAAAAGTTATCTACCCTACCTAACCAAGGTTCGAACCA 102972
Qy 25321 ccttgggtttactgtgagaccacctcctgcagggggcacaactgtcttcagttcagtt 25380
Db 102973 CCTGGGTTTGTGTGAGACCCACCTCCTGCAGGGGCCAACTAGTCTTTCAGTCTCAGTT 103032
Qy 25381 cactggaatttctgagaatccttttagcctggactgctgcacacagtcacagtcattga 25440
Db 103033 CACTTGAATAATTCAGGAATCCTTTTAGCCCTGGACTGCTGCACACAGTCATGGCATTTGA 103092
Qy 25441 gctcagcacagacctgtgagacaggttggctcctctgtgagtgggaaagcagggcctg 25500
Db 103093 GCCTCAGCACAGACCTGTGAGACAGGTGTTGCCCTTGTGAGTGGGAAAGCCAGGCCCTG 103152
Qy 25501 accttggccttccggaatgaaggggcagagcggagccagggcctgttttccagagct 25560
Db 103153 ACCCTTGGCTTTCGGGAATGAAGGGGCAGAGCCGGAGCCGCTCGTTTTTTCAGGAGCT 103212
Qy 25561 tgattttgagagcatctggaactgtctcctctcctcctcctcctccttagccagct 25620
Db 103213 TGATTTTTCAGAGCATCTGGACTGCTCTCCCTTCCCTCTCCGAGGCCCTTAGCCAGCCT 103272
Qy 25621 ggggagcctctgcccctttagagggttcccctccatgcacattcttttccattcagctg 25680
Db 103273 GGGGAGCCTCTGCCCTTTAGAGGGTTCCTCCATGCCATTTCTTTTTCATTTACGCTG 103332
Qy 25681 tggcctgttgctgtgccaagaaaggcgcttggcgctgtgtgtgagcaacatgactgc 25740
Db 103333 TGGCCTGTGGCTTGTGCCAAGNAAGGGCGTTGGCGCTGTGTGTGACACATGACTGC 103392
Qy 25741 atcccttccagctcctgtccccacccttgcccctctgagacatgtccttctctctatt 25800
Db 103393 ATCCCTTCAGCTCTCTGTCGCCCCACCCCTGCCCTCTGAGACATGTCTTGTCTTCTATT 103452
Qy 25801 gtgtcttctaggtctgttggagactttccaaaatgcccaagctgattcccccaactga 25860
Db 103453 GTGTCTTCTAGTCTGCTTTGGAGACTTTTCCAACAATGCCAAGCTGATTCCCCCACTGA 103512
Qy 25861 aagcctgttctgtggaacatcgaggagatggaaggagcagggagaaagcgggagctgc 25920
Db 103513 AAGCCTGTCTGTGGCAGACATCGAGAGATGGAGAGCGCAGGGAGAGAGCGGGAGCTGC 103572
Qy 25921 ggcaggtcggggcgccctcctgtccaggaactggaagcgccctgagggatgggagc 25980
Db 103573 GGCAGTCCGGCGGCCCTCTCTGTCCAGGAGCTGGCAGCGGCCCTGGAGGATGGGGAGC 103632
Qy 25981 ctcacgaaagcggggccacacagagggccacagccaagaggtcagagcccagtgaa 26040
Db 103633 CTCACAGAGGGGGCCACACAGAGAGCCACAGGCCAAGAAGGTCAGAGGCCAGTGAA 103692
Qy 26041 gatctgggagacctgaactcagaaggtgtgtgttcttctcccccaogcaoccgta 26100
Db 103693 GATCTGGGAGACCCCTGAACCTCAGAAGGCTGTGTGTCTTCTGCCCAACGACACGCCGTA 103752

Beck, A., Camp, N.J., Carillo, A.R., Chen, Y., Dayananth, P., Desrochers, M., Dumont, M., Farnham, J.M., Frank, D., Frye, C., Ghaffari, S., Gupte, J.S., Hu, R., Iliev, D., Janecki, T., Kort, E.N., Lait, K.E., Leavitt, A., Leblanc, G., McArthur-Morrison, J., Pederson, A., Penn, B., Peterson, K.T., Reid, J.E., Richards, S., Schroeder, M., Smith, R., Snyder, S.C., Swedlund, B., Swensen, J., Thomas, A., Tranchant, M., Woodland, A.M., Labrie, F., Skolnick, M.H., Neuhausen, S., Rommens, J. and Cannon-Albright, L.A.
A candidate prostate cancer susceptibility gene at chromosome 17p Nat. Genet. 27 (2), 172-180 (2001)
11175785

2 (bases 1 to 2908)
Tavtigian, S.V., Smard, J., Teng, D.H.F., Baumgard, M., Beck, A., Camp, N.J., Carillo, A.R., Chen, Y., Dayananth, P., Desrochers, M., Dumont, M., Farnham, J.M., Frank, D., Frye, C., Ghaffari, S., Gupte, J.S., Hu, R., Iliev, D., Janecki, T., Kort, E.N., Lait, K.E., Leavitt, A., Leblanc, G., McArthur-Morrison, J., Pederson, A., Penn, B., Peterson, K.T., Reid, J.E., Richards, S., Schroeder, M., Smith, R., Snyder, S.C., Swedlund, B., Swensen, J., Thomas, A., Tranchant, M., Woodland, A.M., Labrie, F., Skolnick, M.H., Neuhausen, S., Rommens, J. and Cannon-Albright, L.A.
Direct Submission
Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt Lake City, UT 84108, USA

FEATURES
Location/Qualifiers
1..2908
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p"
1..2481
/codon_start=1
/product="putative prostate cancer susceptibility protein HPC2/ELAC2"
/protein_id="AAC24441.1"
/db_xref="GI:10880933"
/translation="MWAALCLSLRSAAAGRTMSOGRTISQAPARRPRKDPKDLRLHRTRE KRPGSCSGPNTVYLVQVVAAGSRDGAALYVSEFNRYLFNCGEGVQRLMQHEKLKV ARLDNLTFRMWSNVYGLSGMILTTLKTLGPKCVLSGPPQLEKYLEAKIFSGPLKSL IELAVPHGAPSEDEDTMTVYQIPIHSEQRGKHQPQWSPERLSRSPERSSDSSN ENAPHPLPHGVSRRGVDVLAFAFICLHLKRGNFVLVKKAKEMGLPVCVTAIAPIIA AVKDGSITHGREILAEELCTPPDPGAAFFVVECDPESFIQICENATFYQYQKKAQ APALVVHAPASVLDVSRYSQWMERFGPDTQHLVLNENCAVSHNLRSHKIQTQLMLI HPDIFLLTFSRCKBEPTLSVPMVQEGCLLYQLRPREWDARDAITICNPEFTLVLE LQINPQQSQVEYRRSAQDGAPAPAEKRQSPPELIFLGTGSAIPMKIRNYSATLVNPSL DTSLLDCGEGFTLCRHVQDQDVLGTAAAFVSHLHADHTHGLPSILLQRLERAL ASLGLKPLLLVAPNQLKAWLQYHNQCEVLHHISMLPAKCLQGBAELSPAVERL LISLTLCDLEEFQTCVLHRCHAFCAVLTGSKWVYSGDTPMCEALVRMGKDAFL LIHEATLEDGEEAEVEKHTSTTSAISVGMRRNAEFLMNHFSQRYAKVPLFSPNFS EKVGFADHMKVCGDFPTMPKLIPPLKALFAGADIEMEEREKRELRQVRAALLISRE LAGLEBDGEPQOKRAHTEPQAKVRAQ"

BASE COUNT 702 a 793 c 826 g 587 t
ORIGIN

Query Match 2.5%; Score 657; DB 89; Length 2908;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 657; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25810 aggtctgttgagactttccacaatgcccaagctgattcccccaactgaaagccctgt 25869
|||||
Db 2252 AGGTCGTCTTGGAGACTTTCACCAATGCCAAGCTGATTCCCCACTCAAGCCCTGT 2311
|||||

QY 25870 ttgctggcagacatcgagagatagaggagcgcaggagagaagcgggagctcgcgaggtgc 25929
|||||

Db 2312 TTGCTGGCGACATCGAGAGATGAGAGGCGCAGGAGAGACGGGAGCTCGCGCAGGTGC 2371
|||||

QY 25930 gggcgccctcctgtccaggagagctggcagcgccctggaggtgggagacctcagcaga 25989
|||||

Db 2372 GGGCGGCGCTTCTGTCCAGGAGAGCTGGCAGCGCGCCTGGAGAGATGGGAGCCTCAGCAGA 2431
|||||

QY 25990 agcgggccacacagaggagccacagggccaagaggttcagagcccgagtcagatctggga 26049
|||||

```
source
1. .2997
/organism="Homo sapiens"
/db_xref="LocusID:60528"
/db_xref="taxon:9606"
/clone="MGC:4102 IMAGE:2820640"
/tissue_type="Lung, small cell carcinoma"
/clone_lib="NIH_MGC_7"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
61..2541
CDS
/codon_start=1
/product="putative prostate cancer susceptibility protein"
/protein_id="AAH01939.1"
/db_xref="GI:12804973"
/translators="MVALCSLLRSAAAGTMSQGRITISQAPARRPRKDPRLHLRTRE
KRPGSGCGGNTVYLOVVAAGSRDGAALYVSEFNRYLFCNCGEGVQRLMOEHLKV
ARLDNIFLTRMHNSVNGSLGSMILTKETGLPKCVLSGPPQLEKYLBAIKFSGPLKG
IELAVRPHSAPEDEFTWYQIPIHSEQRGRKHPQWSPERLSPUSPERSDSSELN
ENEPHLPHGVSRQRRVDSLLVAFICKLHNRGNFLVAKEMGLPVGTAATAPIA
AVKDKSITHEGREILAEELCTPPDGAFAFVVECPDESFIQPTCENATQRYQGNL
APVALVHMAPASVLDVSRYOQWMEFGPDTHLVNLNENASVHNLRSKTIQTLNLI
HPDIFPLTSTFRCKKEGPTLSVPMVQCECLLYOLRPRRQWDAITTCNPEEFIEA
LQLPNFQSQVQYERSAQDGPAPAEKRSQYPELIIFLGTSAPMKINVSATLVNISP
DTSLLDCGGEGTGQICRHYGDQDVRVLGTLAAVYSHLHADHTGLPSILLQERLAL
ASLGKPLHLPLLVAPNQLKAWLOQYHNQCEVLHHISMIPAKCLQEGAEISSPAVERL
ISSLLRTCDLEEFQTLVCRHKHAFGALVHTSGWKVYSGDTMPCALVRMGKDATL
LIHEATLEDGLEEAVEKTHSTTSQATSVGMNMAEFIMLNHFQRYAKVPLFSPNFS
EKGVAFDHMKVCFGDPPTMKLIPPLKALFAGDIEEMERERREKRLQVRAALLSRE
LAGLEDGEQQRKAHFEPOAKVRQAQ"
740 a 806 c 853 g 598 t
BASE COUNT
ORIGIN

Query Match 2.33; Score 606; DB 91; Length 2997;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 656; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 25810 aggtctgtttggagacttccacaatgcccaagctgattccccactgaaagccctgt 25869
Db 2312 AGGCTGCTTTGGAGACTTCCACAAATGCCCAAGCTGATCCCCCCTGAAGCCCTGT 2371

Qy 25870 ttgtggcgacatcagagagatgaggagcgacgaggaagcggagctgcggcaggtgc 25929
Db 2372 TTGCTGGCGACATCAGGAGATGGAGAGCGACGAGGAGCGGAGCTGCGGAGGTGC 2431

Qy 25930 gggcgccctcctgtccaggagctggcaggcgccctggagagatgggagcctcagcgga 25989
Db 2432 GGGCGGCCCTCCTGTCTCCAGGAGCTGGCAGGCGGCTGGAGGATGGGAGGCTCAGCAGA 2491

Qy 25990 agcgggcccacagagagcagcagggcccaagaaggtcagaccagtcgaagatctcgga 26049
Db 2492 AGCGGGCCCCACAGAGGAGGACACAGGCCAGAGAGCTCAGAGCCAGTGAAGATCTGGGA 2551

Qy 26050 gaccctgaactcagaagctgtgtgtcttctgcccacgcagcaccgctatctgcctc 26109
Db 2552 GACCTGAACCTCAGAGGCTGTGTCTCTTCGCCCCACGACGACCGATCTGCCCTC 2611

Qy 26110 cttgctgtagaagctgaagcagcggtcccccaggagcgagctcagagatgagtggtatg 26169
Db 2612 CTTGCTGTAGAGCTGAAGACAGCGGTCCGCCAGGAGGAGCTCAGGATAGGTGGTATG 2671

Qy 26170 gagctgtccagagcttgggtccacataagcactagctctatagatgcctcttaggact 26229
Db 2672 GAGCTGTCCGAGGCTTGGGCTCCACATAAGCACTAGCTCTATAGATGCCCTCTTAGGACT 2731

Qy 26230 ggtgctggcagcgcgggccagaggtgtgcacacggaagcagcagcagatgaactaat 26289
Db 2732 GGTGCTGGCAGCGCGGGCCAGGAGGTGCCACACGGAAGCAGCAGATGAACATAAT 2791

Qy 26290 ttcatctcaaggcagcttttttaagaagtcttggaaacagcagcgcgccaccttctcttaa 26349
Db 2792 TTCAATTCGAAGGCACTTTTAAAGAAGTCATGGAACAGCAGCGCGGCGCACCTTCTCTTAA 2851
```


RESULT 7

JOURNAL
COMMENT

Unpublished (2000)

Contact: Michael Olivier, David R. Cox
Stanford Human Genome Center
Stanford University School of Medicine
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
Tel: (650) 320-5800
Fax: (650) 320-5801
Email: olivier@shgc.stanford.edu
Primer A: TTTCCTGGATTAGAGGAAGGTG
Primer B: AGTGAAGATCTGGAGACCCTGAA
STS size: 322
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 30 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700

Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Amplifrag Gold Polymerase: 0.07 units/uL
Total Vol: 5 uL

Buffer: MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

Finished human sequence in NCBI. STSS designed and developed at the
Stanford Human Genome Center.

FEATURES
source

1. .429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17"
/clone_lib="Human"

STS

primer_bind 106. .427

primer_bind complement(405. .427)

BASE COUNT 91 a 115 c 108 g 115 t

ORIGIN

Query Match 0.9%; Score 245; DB 54; Length 429;
Best Local Similarity 99.5%; Pred. No. 4.4e-125;
Matches 415; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 26047 ggagacctgaactagaaggctgtgtctcttctgtccacgacgacccctatctgcc 26106

DB 416 GGAGACCCCTGAACCTCAGAAGGCTGTGTCTCTTGTGCCACGACGACCCGTATCTGCC 357

QY 26107 ctctctgtggtagaagctgaagacacgctcccccagagcagctcaggtatggt 26166

DB 356 CTCCTTGCTGTAGAGCTGAAGACGACGCTCCCCAGGAGGAGCTCAGGATAGTGT 297

QY 26167 atggagctgtccgagggcttgggtcccccacataagcactagtctatagctctttag 26226

DB 296 ATGGAGCTGTCCGAGGCTTGGCTGCCACATAGACACTAGTCTATAGATGCTCTTAGG 237

QY 26227 actggtgctggcacagccgagggcagagctgccacacggaagcagagatgaact 26286

DB 236 ACTGTGTGCT-GCACAGCCGCGGGCGAGAGGCTGCCACACGGAAGCAAGCAGATGAAC 178

QY 26287 aatttcattcaaggcaggttttaagaagcttctggaacagcgccgaccttcctc 26346

DB 177 AATTTCAATTCAGGCAGTCTTTTAAGAACTCATGGAACACACGCGGCACCTTCTC 118

QY 26347 taatccagcaagtgattccctgcaccagagacaagcagagatcaacaggtcagtggt 26406

|||||

DB 117 TAATCAGCAAGTGTATCCCTGCACACGACAGACAGCAAGCAAGTAAACGATCAGTGGT 58
QY 26407 ctaagtgtccgagacttaacgaaataagtagtatttcacgtcaataaagattgattg 26463
|||||

DB 57 CTAAGTGTCCGAGACTTAACGAAATAGTATTTCAGCTGCAATAAAGATTGAGTTTG 1

RESULT 10

G58081/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Michael Olivier, David R. Cox

Stanford Human Genome Center

Stanford University School of Medicine

4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA

Tel: (650) 320-5800

Fax: (650) 320-5801

Email: olivier@shgc.stanford.edu

Primer A: CTTATGGAATCTCTCAGCAACACG

Primer B: CTGCACCTGCCTAAACTTTCTGT

STS size: 191

PCR Profile:

Initial incubation: 95 degrees C for 10 minutes

Denaturation: 94 degrees C for 30 seconds

Annealing: 60 degrees C for 30 seconds

Polymerization: 72 degrees C for 23 seconds

PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9700

Protocol:

Template: 25 ng

Primer: each 1 uM

dNTPs: each 200 uM

Amplifrag Gold Polymerase: 0.07 units/uL

Total Vol: 5 uL

Buffer:

MgCl2: 2.5 mM

KCl: 50 mM

Tris-HCl: 10 mM

pH: 8.3

BAC ends sequenced at TIGR from the RPC111 BAC library. Designed

and developed at the Stanford Human Genome Center.

FEATURES

source

1. .220

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="17"

/clone_lib="Human"

2. .192

2. .24

complement(170. .192)

primer_bind 64 a 43 c 67 g 46 t

BASE COUNT

ORIGIN

Query Match 0.8%; Score 210; DB 54; Length 220;

Best Local Similarity 100.0%; Pred. No. 1.7e-105;

Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15459 ccagcgctacaccttctgcacctgcctaaacttctgtgggtattctgcttcccag 15518

LOCUS	AC010458	105787 bp	DNA		PRI	04-OCT-2000
DEFINITION	Homo sapiens chromosome 19 clone CTD-2265M8, complete sequence.					
ACCESSION	AC010458					
VERSION	AC010458.5	GI:10567845				
KEYWORDS	HTG.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 105787)					
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.					
TITLE	Direct Submission					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 105787)					
AUTHORS	DOE Joint Genome Institute.					
TITLE	Direct Submission					
JOURNAL	Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint					
	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA					
REFERENCE	3 (bases 1 to 105787)					
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.					
TITLE	Direct Submission					
JOURNAL	Submitted (04-OCT-2000) DOE Joint Genome Institute, 2800 Mitchell					
	Drive, Walnut creek, CA 94598, USA					
COMMENT	On Oct 4, 2000 this sequence version replaced gi:7711447. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov Finishing Completed at Stanford Human Genome Center www.shgc.stanford.edu Quality: Phrap Quality >=40 99.8% of Sequence; Estimated Total Number of Errors is 0.5. STS Content: SHGC-15011 GI7012.					
FEATURES	Location/Qualifiers					
source	1..105787					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/chromosome="19"					
	/clone="CTD-2265M8"					
BASE COUNT	26919 a 25198 c 25837 g 27833 t					
ORIGIN						
Query Match	0.3%; Score 80; DB 87; Length 105787;					
Best Local Similarity	100.0%; Pred. No. 3.2e-32;					
Matches	80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Qy 17371	gagttcactcttgtgccaggctgagtgcgaatggcgtagtcacgtcacgtcaacc 17430					
	Db 59515 GAGTTTCACTCTTGTCGCCAGGTGGAGTGCATGGCGTAGTCAGTCAACC 59456					
Qy 17431	tccgcctcccgggttcaagc 17450					
Db 59455	TCCGCCTCCC GG GTTCAA GC 59436					
RESULT 15						
AC008158/c						
LOCUS	AC008158	151889 bp	DNA	HTG		09-MAR-2001
DEFINITION	Homo sapiens chromosome 17 clone RP11-42F20 map 17, *** SEQUENCING					
	IN PROGRESS ***, 7 unordered pieces.					
ACCESSION	AC008158					
VERSION	AC008158.12	GI:13123374				
KEYWORDS	HTG; HTGS_PHASEI.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 151889)					
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.					
TITLE	Homo sapiens chromosome 17, clone RP11-42F20					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 151889)					
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.					

Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funker,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Mollia,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tessaye,S., Torruella-Miller,I., Vassiliev,H., Vo.A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

Direct Submission
Submitted (28-JUL-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 25, 2001 this sequence version replaced gi:11181825.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

***** Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu

***** Project Information
Center project name: L552
Center clone name: 42_F-20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 76099: contig of 76099 bp in length
* 76100 76199: gap of 100 bp
* 76200 88646: contig of 12447 bp in length
* 88647 88746: gap of 100 bp
* 88747 95363: contig of 6617 bp in length
* 95364 95463: gap of 100 bp
* 95464 108793: contig of 13330 bp in length
* 108794 108893: gap of 100 bp
* 108894 113782: contig of 4889 bp in length
* 113783 113882: gap of 100 bp
* 113883 130398: contig of 16516 bp in length
* 130399 130498: gap of 100 bp
* 130499 151889: contig of 21391 bp in length.

FEATURES
source

1..151889
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-42F20"
/clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 45858 a 32638 c 31965 g 40824 t 604 others
ORIGIN

Query Match 0.3%; Score 80; DB 60; Length 151889;
Best Local Similarity 100.0%; Pred. No. 3.4e-32;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12635 aattttgtatttttagtagacagcgggtttctccacgttggtcaggctggtctcaaatc 12694
|||||
Db 115986 AATTGTGATTTTATGAGACGGGGTTCTCCACGTTGGTCAGGCTGCTCAAACTC 115927
|||||
QY 12695 ctgacctcaggtgatctgcc 12714
|||||

Db 115926 CTGACCTCAGGTGATCTGCC 115907

RESULT 16
HS407F11

LOCUS
DEFINITION
HS407F11 221507 bp DNA 12-DEC-1999
Human DNA sequence from clone CTA-407F11 on chromosome 22q12
Contains ADRBK2 gene for adrenergic beta receptor kinase 2, the
YESP (v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1)
pseudogene, ESTs, a ca repeat polymorphism, genomic marker D22S421,
STSs, GSSs and two putative CpG islands, complete sequence.

ACCESSION

AL022329

VERSION

AL022329.9

KEYWORDS

HTG; ADRBK2; adrenergic; ca repeat polymorphism; CpG island; human.

SOURCE

Homo sapiens

REFERENCE

1 (bases 1 to 221507)

AUTHORS

Williams,S.

TITLE

Direct Submission

JOURNAL

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

COMMENT

On Jun 7, 1999 this sequence version replaced gi:4688873.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr22>
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPepp; Information
on the WormPepp database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
from the human BAC library described in U-J. Kim et al. (1996)
Genomics 34, 213-218.
VECTOR: pBelOBAC11

This sequence is the entire insert of clone CTA-407F11 The true
right end of clone CTA-390C10 is at 11682 in this sequence. The
start of this sequence overlaps with sequence AL008721 The end of
this sequence overlaps with sequence Z98949.

FEATURES
source

1..221507
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="q12"
/clone="CTA-407F11"
/clone_lib="CIT978SK-A2"
2..415
/note="MERIA repeat: matches 98. .527 of consensus"
repeat_region
repeat_region 424..700
/note="L1ME repeat: matches 5460. .5757 of consensus"
repeat_region 706..853
/note="MIR repeat: matches 52. .212 of consensus"
repeat_region 1167..1210
/note="22 copies 2 mer ta 77 conserved"
repeat_region 1213..1293
/note="MLTIAL repeat: matches 1. .89 of consensus"
repeat_region 1294..1594

```
repeat_region /note="AluY repeat: matches 1. .301 of consensus"
1595. .1918
/note="MT1A1 repeat: matches 89. .364 of consensus"
repeat_region 2014. .2497
/note="MT1E repeat: matches 1. .519 of consensus"
repeat_region 2511. .3278
/note="L1PA10 repeat: matches 5367. .6165 of consensus"
repeat_region 3280. .3345
/note="MT1F repeat: matches 474. .568 of consensus"
repeat_region 3397. .3709
/note="AluJo repeat: matches 1. .312 of consensus"
repeat_region 3744. .4326
/note="MER21B repeat: matches 32. .598 of consensus"
repeat_region 4327. .4597
/note="AluJb repeat: matches 9. .281 of consensus"
repeat_region 4598. .4781
/note="MER21B repeat: matches 598. .790 of consensus"
repeat_region 5622. .6159
/note="L2 repeat: matches 2107. .2747 of consensus"
misc_feature 6815. .7105
/note="match: STS: Em:L02019
match: STS: Em:L02019"
misc_feature 6919. .7264
/note="match: STS: Em:Z23711"
repeat_region 6983. .7034
/note="26 copies 2 mer ca 100 conserved"
repeat_region 7126. .7458
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 7765. .7941
/note="L2 repeat: matches 2290. .2471 of consensus"
repeat_region 8032. .8155
/note="MIR repeat: matches 8. .123 of consensus"
repeat_region 8156. .8466
/note="AluX repeat: matches 1. .311 of consensus"
repeat_region 8467. .8581
/note="MIR repeat: matches 123. .262 of consensus"
repeat_region 8688. .8737
/note="L2 repeat: matches 2700. .2749 of consensus"
repeat_region 8873. .8966
/note="L2 repeat: matches 2020. .2113 of consensus"
repeat_region 9063. .9335
/note="L2 repeat: matches 2030. .2303 of consensus"
repeat_region 9510. .9585
/note="MER5A repeat: matches 32. .109 of consensus"
repeat_region 11072. .11169
/note="MIR repeat: matches 48. .147 of consensus"
misc_feature complement(11107. .11644)
/note="match: GSS: Em:B18104"
misc_feature complement(11252. .11882)
/note="match: GSS: Em:B14354"
repeat_region 11266. .11483
/note="MIR repeat: matches 6. .253 of consensus"
misc_feature 11670. .12192
/note="match: GSS: Em:AQ057571"
misc_feature 11693. .12092
/note="match: GSS: Em:B14426"
repeat_region 12336. .12380
/note="MER5B repeat: matches 74. .127 of consensus"
repeat_region 12945. .12982
/note="MIR repeat: matches 218. .251 of consensus"
repeat_region 12983. .13292
/note="AluYb8 repeat: matches 1. .313 of consensus"
repeat_region 13293. .13545
/note="AluX repeat: matches 2. .245 of consensus"
repeat_region 13546. .13733
/note="MIR repeat: matches 2. .-35 of consensus"
misc_feature complement(14585. .14984)
/note="match: GSS: Em:AQ132582"
misc_feature complement(14742. .14990)
/note="match: GSS: Em:AQ148060"
misc_feature 15262. .16143
/note="CpG island"
/evidence=not_experimental

repeat_region 15291. .15377
/note="29 copies 3 mer gga 77 conserved"
15526. .15725
/note="100 copies 2 mer gg 56 conserved"
15573. .15603
/note="Single clone region"
unsure 15684. .15687
/note="Single clone region"
misc_feature 15688. .15719
/note="weak data"
mRNA join(15726. .15895,55119. .55195,95334. .95407,
112318. .112419,114372. .114446,118481. .118542,
123038. .123089,125179. .125270,129558. .129657,
135912. .135990,138279. .138409,140931. .141025,
145849. .145956,154252. .154318,154851. .154951,
160640. .160706,161810. .161905,165150. .165312,
168987. .169123,172026. .172139,173031. .174689)
/genes="ADRBK2"
/note="match: cDNAs: Em:M73216 Em:M87855 Em:X69117
Em:M34019 Em:M87854 Em:M80776 Em:X61157 Em:S81843
Em:AF087455 Em:AJ223716 Em:L15388 Em:AF074714 Em:D49836
Em:AF135794 Em:AF019766 Em:AF063016
match: ESTs: Em:AI322769 Em:AA312780 Em:AA250907
Em:AA261832 Em:AA369787 Em:AI478542 Em:AI638249
Em:AA579796 Em:AA897081 Em:AA719176 Em:AA250850
Em:AJ568705 Em:AI631522 Em:N20991 Em:N28478 Em:AA278408
Em:AA287692 Em:AA648393 Em:AA279133 Em:AI183614 Em:N25146
Em:T72685 Em:HI5249 Em:AA287659 Em:AI535933 Em:H98627
Em:T72615 Em:AA047782 Em:AA741445 Em:R63891 Em:AA934673
Em:AA057617 Em:T97587 Em:F07794 Em:R63890 Em:AA325870
Em:T29185 Em:R07780 Em:Z45735 Em:C16726 Em:AA322997
Em:T97541 Em:W86958 Em:H55606"
/evidence=not_experimental
/product="br407F11.2 (adrenergic, beta, receptor kinase
2)"
gene join(15726. .15895,55119. .55195,95334. .95407,
112318. .112419,114372. .114446,118481. .118542,
123038. .123089,125179. .125270,129558. .129657,
135912. .135990,138279. .138409,140931. .141025,
145849. .145956,154252. .154318,154851. .154951,
160640. .160706,161810. .161905,165150. .165312,
168987. .169123,172026. .172139,173031. .174689)
/genes="ADRBK2"
CDS join(15783. .15895,55119. .55195,95334. .95407,
112318. .112419,114372. .114446,118481. .118542,
123038. .123089,125179. .125270,129558. .129657,
135912. .135990,138279. .138409,140931. .141025,
145849. .145956,154252. .154318,154851. .154951,
160640. .160706,161810. .161905,165150. .165312,
168987. .169123,172026. .172139,173031. .173192)
/genes="ADRBK2"
/note="match: proteins: Sw:P35626"
/codon_start=1
/evidence=not_experimental
/product="br407F11.2 (adrenergic, beta, receptor kinase
2)"
/protein_id="CAB45657.1"
/db_xref="GI:5139484"
/translacion="MADLEAVLADSVLMMMEKSKATPAARASKRVLPEPSIRSYMQ

Query Match 0.3%; Score 80; DB 92; Length 221507;
Best Local Similarity 100.0%; Pred. No. 3.6e-32;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12635 aatttggatttttagagacggggtttctccacgttggtcagctggtctcaactc 12694
|||||
Db 144699 AATTTTGTATTTTAGTAGAGCGGGGTTTCTCCAGTTGTCAGCTGCTCAAACTC 144758
|||||

QY 12695 ctgacctcagggtatctgcc 12714
|||||
Db 144759 CTGACCTCAGGTATCTGCC 144778
```


REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 121067)
Taudien,S., Bleischmidt,K., Menzel,U., Polley,A., Reichwald,K.,
Rump,A., Schilhabel,M.B., Schudy,A., Wen,G., Siebert,R.,
Schlegelberger,B. and Rosenthal,A.
TITLE Chromosome 8 genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 121067)
AUTHORS Genome Sequencing Center Jena.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-2000) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1132: contig of 1132 bp in length
* 1133 1232: gap of unknown length
* 1233 2191: contig of 959 bp in length
* 2192 2291: gap of unknown length
* 2292 3307: contig of 1016 bp in length
* 3308 3407: gap of unknown length
* 3408 38915: contig of 35408 bp in length
* 38916 38915: gap of unknown length
* 38916 64023: contig of 25108 bp in length
* 64024 64123: gap of unknown length
* 64124 75307: contig of 11184 bp in length
* 75308 75407: gap of unknown length
* 75408 90988: contig of 15581 bp in length
* 90989 91088: gap of unknown length
* 91089 100156: contig of 9068 bp in length
* 100157 100256: gap of unknown length
* 100257 111644: contig of 11388 bp in length
* 111645 111744: gap of unknown length
* 111745 116812: contig of 5068 bp in length
* 116813 116912: gap of unknown length
* 116913 120198: contig of 3286 bp in length
* 120199 120298: gap of unknown length
* 120299 121067: contig of 769 bp in length.

FEATURES
source

1. .121067
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8q13"

/clone="CTB-482E7"

/note="assembly_fragment
vector_side:left"
clone_end:SP6
120299..121067
/note="assembly_fragment
clone_end:T7
vector_side:right"

BASE COUNT 36277 a 24368 c 23800 g 35505 t 1117 others
ORIGIN

Query Match 0.3%; Score 79; DB 78; Length 121067;
Best Local Similarity 100.0%; Pred. No. 1.2e-31;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17555 caggctgggtctgaactcctgacctcaggtgatccaccacccctcagctcccaagtgtt 17614
|||||

Db 73076 CAGGCTGGGTCGAACTCTGACCTCAGGTGATCCACCCACCTTCAGCTCCCAAGTGT 73135
|||||

QY 17615 gggattacaggcgtgagcc 17633
|||||

Db 73136 GGGATTACAGGCGTGAGCC 73154
|||||

RESULT 19
LOCUS AC015663/C
DEFINITION Homo sapiens clone RP11-111L1, *** SEQUENCING IN PROGRESS ***, 38
unordered pieces.

ACCESSION AC015663
VERSION AC015663.3 GI:7523725
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 140801)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome, clone RP11-111L1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 140801)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balgwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Horton,L.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE Direct Submission
JOURNAL Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Apr 7, 2000 this sequence version replaced gi:6524242.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1272
Center clone name: 111_L_1

* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1177: contig of 1177 bp in length
* 1178 1277: gap of 100 bp
* 1278 2445: contig of 1168 bp in length
* 2446 2545: gap of 100 bp
* 2546 3557: contig of 1012 bp in length
* 3558 3657: gap of 100 bp
* 3658 4743: contig of 1086 bp in length
* 4744 4843: gap of 100 bp
* 4844 6600: contig of 1757 bp in length
* 6601 6700: gap of 100 bp
* 6701 8248: contig of 1548 bp in length
* 8249 8348: gap of 100 bp
* 8349 9400: contig of 1052 bp in length
* 9401 9500: gap of 100 bp
* 9501 10588: contig of 1086 bp in length
* 10589 10688: gap of 100 bp


```
* 10689 11717: contig of 1029 bp in length
* 11718 11817: gap of 100 bp
* 11818 13151: contig of 1334 bp in length
* 13152 13251: gap of 100 bp
* 13252 14779: contig of 1528 bp in length
* 14780 14879: gap of 100 bp
* 14880 16587: contig of 1708 bp in length
* 16588 16687: gap of 100 bp
* 16688 18553: contig of 1866 bp in length
* 18554 18653: gap of 100 bp
* 18654 20391: contig of 1738 bp in length
* 20392 20491: gap of 100 bp
* 20492 21670: contig of 1179 bp in length
* 21671 21770: gap of 100 bp
* 21771 24636: contig of 2866 bp in length
* 24637 24736: gap of 100 bp
* 24737 27263: contig of 2527 bp in length
* 27264 27363: gap of 100 bp
* 27364 29840: contig of 2477 bp in length
* 29841 29940: gap of 100 bp
* 29941 31240: contig of 1300 bp in length
* 31241 31340: gap of 100 bp
* 31341 34247: contig of 2907 bp in length
* 34248 34347: gap of 100 bp
* 34348 36876: contig of 2529 bp in length
* 36877 36976: gap of 100 bp
* 36977 38709: contig of 1733 bp in length
* 38710 38809: gap of 100 bp
* 38810 42098: contig of 3289 bp in length
* 42099 42198: gap of 100 bp
* 42199 46578: contig of 4380 bp in length
* 46579 46678: gap of 100 bp
* 46679 50607: contig of 3929 bp in length
* 50608 50707: gap of 100 bp
* 50708 53471: contig of 2764 bp in length
* 53472 53571: gap of 100 bp
* 53572 56915: contig of 3344 bp in length
* 56916 57015: gap of 100 bp
* 57016 60281: contig of 3266 bp in length
* 60282 60381: gap of 100 bp
* 60382 66033: contig of 5652 bp in length
* 66034 66133: gap of 100 bp
* 66134 69824: contig of 3691 bp in length
* 69825 69924: gap of 100 bp
* 69925 75196: contig of 5272 bp in length
* 75197 75296: gap of 100 bp
* 75297 80436: contig of 5140 bp in length
* 80437 80536: gap of 100 bp
* 80537 85035: contig of 4499 bp in length
* 85036 85135: gap of 100 bp
* 85136 92790: contig of 7655 bp in length
* 92791 92890: gap of 100 bp
* 92891 103130: contig of 10240 bp in length
* 103131 103230: gap of 100 bp
* 103231 109728: contig of 6498 bp in length
* 109729 109828: gap of 100 bp
* 109829 125696: contig of 15868 bp in length
* 125697 125796: gap of 100 bp
* 125797 140801: contig of 15005 bp in length.
```

FEATURES source

```
1..140801
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-1111"
/clone_lib="RP11-11 Human Male BAC"
1..1177
/note="assembly_fragment"
1278..2445
/note="assembly_fragment"
2546..3557
/note="assembly_fragment"
3658..4743
/note="assembly_fragment"
```

```
misc_feature 4844..6600
/note="assembly_fragment"
misc_feature 6701..8248
/note="assembly_fragment"
misc_feature 8349..9400
/note="assembly_fragment"
misc_feature 9501..10588
/note="assembly_fragment"
misc_feature 10689..11717
/note="assembly_fragment"
misc_feature 11818..13151
/note="assembly_fragment"
misc_feature 13252..14779
/note="assembly_fragment"
misc_feature 14880..16587
/note="assembly_fragment"
misc_feature 16688..18553
/note="assembly_fragment"
misc_feature 18654..20391
/note="assembly_fragment"
misc_feature 20492..21670
/note="assembly_fragment"
misc_feature 21771..24636
/note="assembly_fragment"
misc_feature 24737..27263
/note="assembly_fragment"
misc_feature 27364..29840
/note="assembly_fragment"
misc_feature 29941..31240
/note="assembly_fragment"
misc_feature 31341..34247
/note="assembly_fragment"
misc_feature 34348..36876
/note="assembly_fragment"
misc_feature 36977..38709
/note="assembly_fragment"
misc_feature 38810..42098
/note="assembly_fragment"
misc_feature 42199..46578
/note="assembly_fragment"
misc_feature 46679..50607
/note="assembly_fragment"
misc_feature 50708..53471
/note="assembly_fragment"
misc_feature 53572..56915
/note="assembly_fragment"
misc_feature 57016..60281
/note="assembly_fragment"
misc_feature 60382..66033
/note="assembly_fragment"
misc_feature 66134..69824
/note="assembly_fragment"
misc_feature 69925..75196
/note="assembly_fragment"
misc_feature 75297..80436
/note="assembly_fragment
clone_end:SP6
vector_side:right"
```

Query Match 0.3%; Score 79; DB 63; Length 140801;
Best Local Similarity 100.0%; Pred. No. 1.2e-31;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17555 caggctggtctcgaaactcctgacctcagtgatccaccacacctcagctcccaagtggt 17614
|||||

Db 73199 CAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCACCTCAGCCTCCCAAGTGTT 73140
|||||

QY 17615 gggattacaggcgtgagcc 17633
|||||

Db 73139 GGGATTACAGCGGTGAGCC 73121
|||||

RESULT 20

```
AC019011 160210 bp DNA HTG 11-APR-2001
LOCUS Homo sapiens chromosome 15 clone RP11-78121 map 15q15, ***
DEFINITION SEQUENCING IN PROGRESS ***, 2 ordered pieces.
AC019011
ACCESSION AC019011.6 GI:13399356
VERSION HTG; HTGS_PHASE2; HTGS_FULLTOP.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 160210)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A.,
Nesbitt,R., Traicoff,R. and Hood,L.
Sequencing of human chromosome 15 D15S146-D15S117 region
Unpublished
2 (bases 1 to 160210)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
and Hood,L.
Direct Submission
Submitted (28-DEC-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
On Mar 21, 2001 this sequence version replaced gi:13310883.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMWSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leetowen@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Insert size: 155000; agarose-fp
Quality coverage: 10.9x in Q20 bases; sum-of-contigs
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 129675: contig of 129675 bp in length
* 129676 129775: gap of unknown length
* 129776 160210: contig of 30435 bp in length.
Location/Qualifiers
1. 160210
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q15"
/clone="RP11-78121"
/clone.lib="RPC1 human BAC library 11"
/note="This clone overlaps RP11-355D13 and RP11-315A19"
47748 a 36665 c 33372 g 42324 t 101 others
BASE COUNT
ORIGIN
```

```
Query Match 0.3%; Score 79; DB 65; Length 160210;
Best Local Similarity 100.0%; Pred. No. 1.2e-31;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17555 caggctggtctgaactcctgcactcaggtgatccacccacctcagctcccaagtgtt 17614
|||||
Db 95877 CAGGCTGGTCTCGAACTCTGACCTCAGGTGATCCACCTCAGCTCCCAAGTGT 95936

QY 17615 gggattaccaggcgtgagcc 17633
|||||
Db 95937 GGGATTACAGCGGTGAGCC 95955

RESULT 21
AC015664 160990 bp DNA HTG 26-FEB-2001
LOCUS Homo sapiens chromosome 15 clone RP11-114F23 map 15, WORKING DRAFT
DEFINITION SEQUENCE, 25 unordered pieces.
AC015664
ACCESSION AC015664
VERSION AC015664.5 GI:13123268
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 160990)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Unpublished
2 (bases 1 to 160990)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhaltier,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donellan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 25, 2001 this sequence version replaced gi:12232534.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: Li273
Center clone name: 114_F-23
----- Summary Statistics
Sequencing vector: M13; M7815; 10% of reads
Chemistry: Dye-terminator Big Dye; 95% of reads
Chemistry: Dye-primer-amersham; 5% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 149544 bases at least Q40
Consensus quality: 155176 bases at least Q30
Consensus quality: 157159 bases at least Q20
Insert size: 165000; agarose-fp
Insert size: 158590; sum-of-contigs
Quality coverage: 4.7 in Q20 bases.
* NOTE: This is a 'working draft' sequence. It currently
```

* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 180: contig of 180 bp in length
* 181 280: gap of 100 bp
* 281 1445: contig of 1165 bp in length
* 1446 1545: gap of 100 bp
* 1546 3093: contig of 1548 bp in length
* 3094 3193: gap of 100 bp
* 3194 5034: contig of 1841 bp in length
* 5035 5134: gap of 100 bp
* 5135 6930: contig of 1796 bp in length
* 6931 7030: gap of 100 bp
* 7031 9142: contig of 2112 bp in length
* 9143 9242: gap of 100 bp
* 9243 11655: contig of 2413 bp in length
* 11656 11755: gap of 100 bp
* 11756 14321: contig of 2566 bp in length
* 14322 14421: gap of 100 bp
* 14422 17472: contig of 3051 bp in length
* 17473 17572: gap of 100 bp
* 17573 20422: contig of 2850 bp in length
* 20423 20522: gap of 100 bp
* 20523 23318: contig of 2796 bp in length
* 23319 23418: gap of 100 bp
* 23419 52902: contig of 29484 bp in length
* 52903 53002: gap of 100 bp
* 53003 55126: contig of 2124 bp in length
* 55127 55226: gap of 100 bp
* 55227 59172: contig of 3946 bp in length
* 59173 59272: gap of 100 bp
* 59273 63905: contig of 4633 bp in length
* 63906 64005: gap of 100 bp
* 64006 69074: contig of 5069 bp in length
* 69075 69174: gap of 100 bp
* 69175 74111: contig of 4937 bp in length
* 74112 74211: gap of 100 bp
* 74212 81143: contig of 6932 bp in length
* 81144 81243: gap of 100 bp
* 81244 88688: contig of 7445 bp in length
* 88689 88788: gap of 100 bp
* 88789 101600: contig of 12812 bp in length
* 101601 101700: gap of 100 bp
* 101701 115111: contig of 13411 bp in length
* 115112 115211: gap of 100 bp
* 115212 131102: contig of 15891 bp in length
* 131103 131202: gap of 100 bp
* 131203 143011: contig of 11809 bp in length
* 143012 143111: gap of 100 bp
* 143112 158767: contig of 15656 bp in length
* 158768 158867: gap of 100 bp
* 158868 160990: contig of 2123 bp in length.

FEATURES

source
1..160990
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-114F23"
/clone_lib="RPC1-11 Human Male BAC"
1..180
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature
281..1445
/note="assembly_fragment"
1546..3093
/note="assembly_fragment"
3194..5034

misc_feature
5135..6930
/note="assembly_fragment"
misc_feature
7031..9142
/note="assembly_fragment"
misc_feature
9243..11655
/note="assembly_fragment"
misc_feature
11756..14321
/note="assembly_fragment"
misc_feature
14422..17472
/note="assembly_fragment"
misc_feature
17573..20422
/note="assembly_fragment"
misc_feature
20523..23318
/note="assembly_fragment"
misc_feature
23419..52902
/note="assembly_fragment"
misc_feature
53003..55126
/note="assembly_fragment"
misc_feature
55227..59172
/note="assembly_fragment"
misc_feature
59273..63905
/note="assembly_fragment"
misc_feature
64006..69074
/note="assembly_fragment"
misc_feature
69175..74111
/note="assembly_fragment"
misc_feature
74212..81143
/note="assembly_fragment"
misc_feature
81244..88688
/note="assembly_fragment"
misc_feature
88789..101600
/note="assembly_fragment"
misc_feature
101701..115111
/note="assembly_fragment"
misc_feature
115212..131102
/note="assembly_fragment"
misc_feature
131203..143011
/note="assembly_fragment"
misc_feature
143112..158767
/note="assembly_fragment"
misc_feature
158868..160990
/note="assembly_fragment"
clone_end:T7
vector_side:right"
BASE COUNT 44409 a 34585 c 34521 g 45055 t 2420 others
ORIGIN

Query Match 0.3%; Score 79; DB 63; Length 160990;
Best Local Similarity 100.0%; Pred. No. 1.2e-31;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 17555 caggctggtctgaactcctgacctcagtgatccaccacacccagctcccaagtggt 17614
|||||
Db 147561 CAGGCTGCTCGAAGTCTCTGACCTCAGGTGATCCACCACCTCCCAAGTGT 147620
|||||
Qy 17615 gggattacaggcgtgagcc 17633
|||||
Db 147621 GGGATTACAGGCGTGAGCC 147639

RESULT 22

AC023829
LOCUS Homo sapiens chromosome 16 clone RP11-578P21, WORKING DRAFT
DEFINITION SEQUENCE, 17 unordered pieces.
ACCESSION AC023829
VERSION AC023829.3 GI:9100196
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 164353)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
DOE Joint Genome Institute.
2 (bases 1 to 164353)
Direct Submission
Submitted (18-FEB-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 13, 2000 this sequence version replaced gi:7211924.

Query Match 0.3%; Score 79; DB 68; Length 164353;
Best Local Similarity 100.0%; Pred. No. 1.2e-31;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17555 caggctggtctgcgaactcctgcactcaggtgatccacccacccagctcccaagtgtt 17614
|||||
DB 24628 CAGGCTGGTCTCGAACTCTCGACTCAGGTGATCCACCCACCTCAGCTCCCAAGTGT 24687
|||||
QY 17615 gggattacagggcgtgagcc 17633
|||||
DB 24688 GGGATTACAGGCGTGAGCC 24706
|||||

Project Information
Center Project Name: 0
Center clone name: RPCI-11_578P21

Summary Statistics
Consensus quality: 120365 bases at least Q40
Consensus quality: 138510 bases at least Q30
Consensus quality: 148253 bases at least Q20
Estimated insert size: 149000; agarose-fp estimation
Estimated insert size: 162753; sum-of-contigs estimation
Quality coverage: 4.29 in Q20 bases; agarose-fp estimation
Quality coverage: 3.93 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

RESULT 23
AC018844/C
LOCUS
DEFINITION
AC018844
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC018844 165173 bp DNA HTG 11-AUG-2000
Homo sapiens chromosome 16 clone RP11-167J20, WORKING DRAFT
SEQUENCE, 3 ordered pieces.
AC018844
AC018844.2 GI:8576104
HTG; HTGS_PHASE2; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 165173)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
DOE Joint Genome Institute.
2 (bases 1 to 165173)
Direct Submission
Submitted (21-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jun 21, 2000 this sequence version replaced gi:6623970.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 37379: contig of 37379 bp in length
* 37380 122440: contig of unknown length
* 37480 122440: contig of 84961 bp in length
* 122441 122540: gap of unknown length
* 122541 165173: contig of 42633 bp in length.
* Location/Qualifiers
1. .165173
source

1 1995: contig of 1995 bp in length
* 1996 2095: gap of unknown length
* 2096 3988: contig of 1893 bp in length
* 3989 4088: gap of unknown length
* 4089 5674: contig of 1586 bp in length
* 5675 5774: gap of unknown length
* 5775 8312: contig of 2538 bp in length
* 8313 8412: gap of unknown length
* 8413 10402: contig of 1990 bp in length
* 10403 10502: gap of unknown length
* 10503 12317: contig of 1715 bp in length
* 12318 12317: gap of unknown length
* 12318 15596: contig of 3279 bp in length
* 15597 15696: gap of unknown length
* 15697 18245: contig of 2549 bp in length
* 18246 18345: gap of unknown length
* 18346 21333: contig of 2988 bp in length
* 21334 21433: gap of unknown length
* 21434 29820: contig of 8387 bp in length
* 29821 29920: gap of unknown length
* 29921 38383: contig of 8463 bp in length
* 38384 38483: gap of unknown length
* 38484 47716: contig of 9233 bp in length
* 47717 47816: gap of unknown length
* 47817 57354: contig of 9538 bp in length
* 57355 57454: gap of unknown length
* 57455 68519: contig of 11065 bp in length
* 68520 68619: gap of unknown length
* 68620 93877: contig of 25258 bp in length
* 93878 93977: gap of unknown length
* 93978 128937: contig of 34960 bp in length
* 128938 129037: gap of unknown length
* 129038 164353: contig of 35316 bp in length.
Location/Qualifiers
1. .164353
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-167J20"

BASE COUNT 50235 a 35260 c 34381 g 45089 t 208 others
ORIGIN

Query Match 0.3%; Score 79; DB 65; Length 165173;
Best Local Similarity 100.0%; Pred. No. 1.2e-31;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17555 caggctggtctgaactcgtacgtcagctcaccacccacgtcctcccaagtggt 17614
|||||
Db 46174 CAGGCTGGTCTCGAACTCTGACCTGAGTGATCCACCACCTCAGCTCCCAAAGTGT 46115
|||||

QY 17615 gggattacaggcgtgagcc 17633
|||||
Db 46114 GGGATTACAGGCGTGAGCC 46096
|||||

RESULT 24
AC018924 182608 bp DNA HTG 14-APR-2001
LOCUS Homo sapiens chromosome 15 clone RP11-355D13 map 15q15, ***
SEQUENCING IN PROGRESS ***, 3 ordered pieces.
ACCESSION AC018924
VERSION AC018924.6 GI:13624385
KEYWORDS HTG; HTGS_PHASE2; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 182608)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
Sequencing of human chromosome 15 D15S146-D15S117 region
Unpublished
2 (bases 1 to 182608)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
and Hood,L.
Direct Submission
Submitted (23-DEC-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
On Apr 14, 2001 this sequence version replaced gi:13399355.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UWMSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Insert size: 180000; agarose-fp
Quality coverage: 9.6x in Q20 bases; sum-of-contigs
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs

```

```

* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 25322: contig of 25322 bp in length
* 25323 25422: gap of unknown length
* 25423 27704: contig of 2282 bp in length
* 27705 27804: gap of unknown length
* 27805 182608: contig of 154804 bp in length.
FEATURES
Location/Qualifiers
1..182608
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q15"
/clone="RP11-355D13"
/clone_lib="RPCI human BAC library 11"
/note="This clone overlaps RP11-402F9 and RP11-78121"

BASE COUNT 54890 a 39492 c 36932 g 51036 t 258 others
ORIGIN

Query Match 0.3%; Score 79; DB 65; Length 182608;
Best Local Similarity 100.0%; Pred. No. 1.3e-31;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17555 caggctggtctgaactcgtcagctcagctcaccacccacgtcctcccaagtggt 17614
|||||
Db 177519 CAGGCTGGTCTCGAACTCTGACCTGAGTGATCCACCACCTCAGCTCCCAAAGTGT 177578
|||||

QY 17615 gggattacaggcgtgagcc 17633
|||||
Db 177579 GGGATTACAGGCGTGAGCC 177597
|||||

RESULT 25
AC078861/c 186892 bp DNA HTG 07-JAN-2001
LOCUS Homo sapiens chromosome 12q clone RP11-1064P9, *** SEQUENCING IN
PROGRESS ***, 36 unordered pieces.
ACCESSION AC078861
VERSION AC078861.13 GI:12039133
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 186892)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,K., Jolivet,S.,
Joudan,S., Kratovick,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovick,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

```



```
|||||GGGATTACAGCGTGAGCC 42697

Db 42715

RESULT 26
G42927/c 137 bp mRNA STS 27-JAN-1999
LOCUS W1AF-176-STS Human Thudson EST Homo sapiens STS cDNA, sequence
DEFINITION tagged site.
ACCESSION G42927
VERSION G42927.1 GI:4191844
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 137)
AUTHORS Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,P., Sapolsky,R.,
Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L.,
Stein,L., Hsieh,L., Topaloglou,T., Hubbell,E., Robinson,E.,
Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J.,
Nusbaum,C., Rozen,S., Hudson,T.J., Lipshutz,R., Chee,M. and
Lander,E.S.
TITLE Large-scale identification, mapping, and genotyping of
single-nucleotide polymorphisms in the human genome
JOURNAL Science 280 (5366), 1077-1082 (1998)
MEDLINE 98248615
COMMENT Synonyms: EST226740b, EST226740
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: TAATGCGAGCTGAATATCTATTTCG
Primer B: AAGTCATGGAACACAGCGC
STS size: 137
PCR Profile:
Presoak: 94 degrees C for 4.00 minutes
Denaturation: 94 degrees C for 50.0 seconds
Annealing: 58 degrees C for 1.50 minutes
Polymerization: 72 degrees C for 1.00 minutes
PCR Cycles: 30
Thermal Cycler: custom built by IAS, Costar, Cambridge MA

Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: 4 nM
Taq Polymerase: 0.5 U
Total Vol: 20 uL

Buffer:
Mg2+: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
Gelatin: .001 %
Location/Qualifiers
1..137
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="36.40 cR from top of Chr17 linkage group"
/clone_lib="Human Thudson EST"
/note="STS derived from sequences in dbEST and the
Unigene collection."
1..137
1..25
STS primer_bind complement(118..137)
primer_bind 28 a 30 c 31 g 46 t 2 others
BASE COUNT 0.38; Score 78; DB 54; Length 137;
ORIGIN

FEATURES
source
1..100272
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP5-842G6"
/clone_lib="RPCI-5"
repeat_region 533..676
/note="MER63A repeat: matches 57..201 of consensus"
repeat_region 835..922
/note="MER63A repeat: matches 109..210 of consensus"
mRNA join(<1831..1878,3489..3592,7112..7151,10137..10334,
17493..17553,25114..25197,25526..25608,25769..27072)

Best Local Similarity 100.0%; Pred. No. 1.7e-31;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26373 accagagacagcagagtaaacaggtatcagtggtggttaagtcagagacgttaacgaaaaat 26432
|||||
Db 78 ACCAGAGACAAGCAGAGTAACAGGATCAGTGGGTCTAAGTGTCCGAGACTTAACGAAAAT 19
QY 26433 agtatttcagctgcaata 26450
|||||
Db 18 AGTATTTTCAGCTGCARTA 1

RESULT 27
HSDJ842G6/c 100272 bp DNA PRI 04-APR-2001
LOCUS Human DNA sequence from clone RP5-842G6 on chromosome 20. Contains
DEFINITION the 3' end of a novel gene, the 3' end of the gene for a novel
protein similar to SEL1L (sel-1 (suppressor of lin-12,
C.elegans)-like), ESTs, STSs and GSSs, complete sequence.
ACCESSION AL109657
VERSION AL109657.8 GI:6136991
KEYWORDS HTG; SEL1L.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 100272)
AUTHORS Barlow,K.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Oct 27, 1999 this sequence version replaced gi:6015555.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence is
the entire insert of clone RP5-842G6 This sequence was generated
from part of bacterial clone contigs of human chromosome 20,
constructed by the Sanger Centre Chromosome 20 Mapping Group.
Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP5-842G6 is from the
library RPCI-5 constructed by the group of Pieter de Jong. For
further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.

FEATURES
source
1..100272
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP5-842G6"
/clone_lib="RPCI-5"
repeat_region 533..676
/note="MER63A repeat: matches 57..201 of consensus"
repeat_region 835..922
/note="MER63A repeat: matches 109..210 of consensus"
mRNA join(<1831..1878,3489..3592,7112..7151,10137..10334,
17493..17553,25114..25197,25526..25608,25769..27072)
```


Qy 17374 ttactcttggccagggtgagtgcaatggcggtgatctcagtcagtcagtcacccctcc 17433
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 11312 TTTCACTCTTGTGCTCCAGGCTGGAGTCAATGGGTGATCTCAGTCTCAGTCCACCTCC 11253
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 17434 gcctcccggtttcaagca 17451
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 11252 GCCTCCCGGTTCAAGCA 11235
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 28
 AC024993
 LOCUS Homo sapiens chromosome 8 clone RP11-262B3 map 8, WORKING DRAFT
 DEFINITION SEQUENCE, 2 ordered pieces.
 AC024993
 AC024993.4 GI:12061519
 VERSION HTG: HTGS_PHASE2; HTGS_DRAFT.
 KEYWORDS human.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 141708)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 8, clone RP11-262B3
 Unpublished
 2 (bases 1 to 141708)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
 Campolano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collumore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczkzy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisan,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (03-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jan 10, 2001 this sequence version replaced gi:8076998.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WBIR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L7450
 Center clone name: 262_B3
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 69% of reads
 Sequencing vector: Plasmid; n/a; 31% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 141185 bases at least Q40
 Consensus quality: 141462 bases at least Q30
 Consensus quality: 141576 bases at least Q20
 Insert size: 135000; agarose-fp
 Insert size: 141608; sum-of-ctnigs
 Quality coverage: 13.1 in Q20 bases; agarose-fp

Quality coverage: 12.5 in Q20.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 2032: contig of 2032 bp in length
 * 2033 2132: gap of 100 bp
 * 2133 141708: contig of 139576 bp in length.

FEATURES
 source
 1. 141708
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-262B3"
 /clone.lib="RPCI-11 Human Male BAC"
 1..2032
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"
 2133..141708
 /note="assembly_fragment
 clone_end:Y7
 vector_side:right"
 BASE COUNT 38281 a 28210 c 29655 g 45462 t 100 others
 ORIGIN

Query Match 0.3%; Score 78; DB 69; Length 141708;
 Best Local Similarity 100.0%; Pred. No. 4.4e-31;
 Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 17556 agcgtggtcgaactcgcacccaggtgacccacccagtcacccacccagtcg 17615
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 15816 AGCGTGGTCTCGAATCTCGAGTCAGGTGATCCACCCACCTCAGCCTCCCAAGGTGTG 15875
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 17616 ggattacagcgctgagcc 17633
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 15876 GGATTACAGCGCTGAGCC 15893
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 29
 AC015520
 LOCUS Homo sapiens clone RP11-23C19, WORKING DRAFT SEQUENCE, 41 unordered
 DEFINITION pieces.
 AC015520
 AC015520.3 GI:10045465
 VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS human.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 180052)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens, clone RP11-23C19
 Unpublished
 2 (bases 1 to 180052)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collumore,A.,
 Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lehoczkzy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE JOURNAL COMMENT

Direct Submission
 Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 9, 2000 this sequence version replaced gi:6642701.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information

Center project name: L4175
 Center clone name: 23_C_19
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 153771 bases at least Q40
 Consensus quality: 165735 bases at least Q30
 Consensus quality: 170554 bases at least Q20
 Insert size: 156000; agarose-fp
 Insert size: 176052; sum-of-contigs
 Quality coverage: 3.9 in Q20 bases; agarose-fp
 Quality coverage: 3.5 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 41 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1528: contig of 1528 bp in length
 1529 1628: gap of 100 bp
 1629 2636: contig of 1008 bp in length
 2637 2736: gap of 100 bp
 2737 3786: contig of 1050 bp in length
 3787 3886: gap of 100 bp
 3887 4904: contig of 1018 bp in length
 4905 5004: gap of 100 bp
 5005 6019: contig of 1015 bp in length
 6020 6119: gap of 100 bp
 6120 7186: contig of 1067 bp in length
 7187 7286: gap of 100 bp
 7287 8382: contig of 1096 bp in length
 8383 8482: gap of 100 bp
 8483 9493: contig of 1011 bp in length
 9494 9593: gap of 100 bp
 9594 10732: contig of 1139 bp in length
 10733 10832: gap of 100 bp
 10833 11904: contig of 1072 bp in length
 11905 12004: gap of 100 bp
 12005 13026: contig of 1022 bp in length
 13027 13126: gap of 100 bp
 13127 14157: contig of 1031 bp in length
 14158 14257: gap of 100 bp
 14258 15325: contig of 1068 bp in length
 15326 15425: gap of 100 bp
 15426 16680: contig of 1255 bp in length
 16681 16780: gap of 100 bp
 16781 17795: contig of 1015 bp in length
 17796 17895: gap of 100 bp
 17896 20340: contig of 2445 bp in length
 20341 20440: gap of 100 bp
 20441 21690: contig of 1250 bp in length
 21691 21790: gap of 100 bp

21791 23814: contig of 2024 bp in length
 23815 23914: gap of 100 bp
 23915 26113: contig of 2199 bp in length
 26114 26213: gap of 100 bp
 26214 27789: contig of 1576 bp in length
 27790 27889: gap of 100 bp
 27890 29265: contig of 1376 bp in length
 29266 29365: gap of 100 bp
 29366 32181: contig of 2816 bp in length
 32182 32281: gap of 100 bp
 32282 33701: contig of 1420 bp in length
 33702 33801: gap of 100 bp
 33802 35279: contig of 1478 bp in length
 35280 35379: gap of 100 bp
 35380 37283: contig of 1904 bp in length
 37284 37383: gap of 100 bp
 37384 40142: contig of 2759 bp in length
 40143 40242: gap of 100 bp
 40243 43686: contig of 3444 bp in length
 43687 43786: gap of 100 bp
 43787 47088: contig of 3302 bp in length
 47089 47188: gap of 100 bp
 47189 52035: contig of 4847 bp in length
 52036 52135: gap of 100 bp
 52136 55272: contig of 3137 bp in length
 55273 55372: gap of 100 bp
 55373 61679: contig of 6307 bp in length
 61680 61779: gap of 100 bp
 61780 67744: contig of 5965 bp in length
 67745 67844: gap of 100 bp
 67845 75958: contig of 8014 bp in length
 75959 93357: contig of 17399 bp in length
 93358 93457: gap of 100 bp
 93458 102195: contig of 8738 bp in length
 102196 102295: gap of 100 bp
 102296 115757: contig of 13462 bp in length
 115758 115857: gap of 100 bp
 115858 130749: contig of 14892 bp in length
 130750 130849: gap of 100 bp
 130850 140678: contig of 9829 bp in length
 140679 140778: gap of 100 bp
 140779 154853: contig of 14075 bp in length
 154854 154953: gap of 100 bp
 154954 173134: contig of 18181 bp in length
 173135 173234: gap of 100 bp
 173235 180052: contig of 6818 bp in length.

FEATURES

Source
 1. .180052
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-23C19"
 /clone_lib="RPC1-11 Human Male BAC"
 misc_feature
 1. .1528
 /note="assembly_fragment
 clone_end:sp6
 vector_side:left"
 misc_feature
 1629. .2636
 /note="assembly_fragment"
 misc_feature
 2737. .3786
 /note="assembly_fragment"
 misc_feature
 3887. .4904
 /note="assembly_fragment"
 misc_feature
 5005. .6019
 /note="assembly_fragment"
 misc_feature
 6120. .7186
 /note="assembly_fragment"
 misc_feature
 7287. .8382
 /note="assembly_fragment"
 misc_feature
 8483. .9493
 /note="assembly_fragment"
 misc_feature
 9594. .10732
 /note="assembly_fragment"

```
misc_feature 10833..11904
/note="assembly_fragment"
misc_feature 12005..13026
/note="assembly_fragment"
misc_feature 13127..14157
/note="assembly_fragment"
misc_feature 14258..15325
/note="assembly_fragment"
misc_feature 15426..16680
/note="assembly_fragment"
misc_feature 16781..17795
/note="assembly_fragment"
misc_feature 17896..20340
/note="assembly_fragment"
misc_feature 20441..21690
/note="assembly_fragment"
misc_feature 21791..23814
/note="assembly_fragment"
misc_feature 23915..26113
/note="assembly_fragment"
misc_feature 26214..27789
/note="assembly_fragment"
misc_feature 27890..29265
/note="assembly_fragment"
misc_feature 29366..32181
/note="assembly_fragment"
misc_feature 32282..33701
/note="assembly_fragment"
misc_feature 33802..35279

Query Match 0.3%; Score 78; DB 63; Length 180052;
Best Local Similarity 100.0%; Pred. No. 4.6e-31;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17556 aggcgtgtctcgaaactctgcacctgagtgatccaccacccatcgctcccccagaagtgttg 17615
|||||
Db 5234 AGGCTGGTCTCGAACTCTGCACCTGAGTGATCCACCACCTCAGCTCCCAAGTGTG 5293

Qy 17616 ggattacaggctgagcc 17633
|||||
Db 5294 GGATTACAGCGGTGAGCC 5311

RESULT 30
AC015465 213943 bp DNA HTG 10-SEP-2000
LOCUS Homo sapiens chromosome 8 clone RP11-35A5 map 8, WORKING DRAFT
DEFINITION SEQUENCE, 31 unordered pieces.
ACCESSION AC015465
VERSION AC015465.4 GI:10047715
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 213943)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
TITLE Homo sapiens chromosome 8, clone RP11-35A5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 213943)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boquslavkiy,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Doneilan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrison,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
```

Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 10, 2000 this sequence version replaced gi:7341762.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2255
Center clone name: 35_A_5
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 192210 bases at least Q40
Consensus quality: 203227 bases at least Q30
Consensus quality: 207438 bases at least Q20
Insert size: 17700; agarose-fp
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 3.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2202: contig of 2202 bp in length
* 2203 2302: gap of 100 bp
* 2303 3678: contig of 1376 bp in length
* 3679 3778: gap of 100 bp
* 3779 5823: contig of 2045 bp in length
* 5824 5923: gap of 100 bp
* 5924 8852: contig of 2929 bp in length
* 8853 8952: gap of 100 bp
* 8953 12186: contig of 3234 bp in length
* 12187 12286: gap of 100 bp
* 12287 16055: contig of 3769 bp in length
* 16056 16155: gap of 100 bp
* 16156 19742: contig of 3587 bp in length
* 19743 19842: gap of 100 bp
* 19843 23449: contig of 3607 bp in length
* 23450 23549: gap of 100 bp
* 23550 27145: contig of 3596 bp in length
* 27146 27245: gap of 100 bp
* 27246 31030: contig of 3785 bp in length
* 31031 31130: gap of 100 bp
* 31131 34297: contig of 3167 bp in length
* 34298 34397: gap of 100 bp
* 34398 38290: contig of 3893 bp in length
* 38291 38390: gap of 100 bp
* 38391 41958: contig of 3568 bp in length
* 41959 42058: gap of 100 bp
* 42059 43381: contig of 3323 bp in length
* 43382 45481: gap of 100 bp
* 45482 53017: contig of 7536 bp in length
* 53018 53117: gap of 100 bp
* 53118 57744: contig of 4627 bp in length
* 57745 57844: gap of 100 bp
* 57845 63406: contig of 5562 bp in length
* 63407 63506: gap of 100 bp
* 63507 83520: contig of 20014 bp in length
* 83521 83620: gap of 100 bp
* 83621 89997: contig of 6377 bp in length

TITLE
JOURNAL
COMMENT

```
* 89998 90097: gap of 100 bp
* 90098 96192: contig of 6095 bp in length
* 96193 96292: gap of 100 bp
* 96293 104654: contig of 8362 bp in length
* 104655 104754: gap of 100 bp
* 104755 110937: contig of 6183 bp in length
* 110938 111037: gap of 100 bp
* 111038 117971: contig of 6934 bp in length
* 117972 118071: gap of 100 bp
* 118072 128570: contig of 10499 bp in length
* 128571 128670: gap of 100 bp
* 128671 138597: contig of 9927 bp in length
* 138598 138697: gap of 100 bp
* 138698 146653: contig of 7956 bp in length
* 146654 146753: gap of 100 bp
* 146754 159779: contig of 13026 bp in length
* 159780 159879: gap of 100 bp
* 159880 179102: contig of 19223 bp in length
* 179103 179202: gap of 100 bp
* 179203 193866: contig of 14664 bp in length
* 193867 193966: gap of 100 bp
* 193967 209155: contig of 15189 bp in length
* 209156 209255: gap of 100 bp
* 209256 213943: contig of 4688 bp in length.
FEATURES
    source
        1..213943
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="8"
            /map="8"
            /clone_lib="RPCI-11 Human Male BAC"
            1..2202
                /note="assembly_fragment
                clone_end:SP6
                vector_side:left"
                2303..3678
                    /note="assembly_fragment"
                3779..5823
                    /note="assembly_fragment"
                5924..8852
                    /note="assembly_fragment"
                8953..12186
                    /note="assembly_fragment"
                12287..16055
                    /note="assembly_fragment"
                16156..19742
                    /note="assembly_fragment"
                19843..23449
                    /note="assembly_fragment"
                23550..27145
                    /note="assembly_fragment"
                27246..31030
                    /note="assembly_fragment"
                31131..34297
                    /note="assembly_fragment"
                34398..38290
                    /note="assembly_fragment"
                38391..41598
                    /note="assembly_fragment"
                42059..45381
                    /note="assembly_fragment"
                45482..53017
                    /note="assembly_fragment"
                53118..57744
                    /note="assembly_fragment"
                57845..63406
                    /note="assembly_fragment"
                63507..83520
                    /note="assembly_fragment"
                83621..89997
                    /note="assembly_fragment"
                90098..96192
```

```
misc_feature
    96293..104654
        /note="assembly_fragment"
misc_feature
    104755..110937
        /note="assembly_fragment"
misc_feature
    111038..117971
        /note="assembly_fragment"
misc_feature
    118072..128570
        /note="assembly_fragment"
misc_feature
    128671..138597
        /note="assembly_fragment"
misc_feature
    138698..146653
        /note="assembly_fragment"
misc_feature
    146754..159779
        /note="assembly_fragment"
misc_feature
    159880..179102
        /note="assembly_fragment"
misc_feature
    179203..193866
        /note="assembly_fragment"
misc_feature
    193967..209155
        /note="assembly_fragment"
misc_feature
    209256..213943
        /note="assembly_fragment
        clone_end:T7
        vector_side:right"
BASE COUNT 62397 a 42447 c 43297 g 62791 t 3011 others
Query Match 0.3%; Score 78; DB 63; Length 213943;
Best Local Similarity 100.0%; Pred. No. 4.7e-31;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17556 agcgtggtcgaactcctgacctgacctgagtgatccaccacccagctcccaagtggtg 17615
|||||
Db 211202 AGCTGGTCTCGAACTCTCTGACCTGAGTGTGATCCACCTCAGCTCCCAAGTGTG 211261
QY 17616 ggattacaggcgtgagcc 17633
|||||
Db 211262 GGATTACAGGCGTGAGCC 211279
RESULT 31
AL136969
LOCUS
DEFINITION
    AL136969 79666 bp DNA PRI 11-APR-2001
    Human DNA sequence from clone RPI-207J11 on chromosome 7 Contains
    ESTs and STSS. Contains two zinc finger protein pseudogenes, a
    putative novel gene and an HNRPC (heterogeneous nuclear
    ribonucleoprotein C (Cl/C2)) pseudogene, complete sequence.
ACCESSION
    AL136969
VERSION
    AL136969.7 GI:10443366
KEYWORDS
    HTG; HNRPC; ribonucleoprotein; zinc finger.
SOURCE
    human.
ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
    1 (bases 1 to 79666)
AUTHORS
    Brown,J.
JOURNAL
    Direct Submission
    Submitted (30-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
    CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
    requests: clonerequest@sanger.ac.uk
    On Oct 1, 2000 this sequence version replaced gi:8247074.
    During sequence assembly data is compared from overlapping clones.
    Where differences are found these are annotated as variations
    together with a note of the overlapping clone name. Note that the
    variation annotation may not be found in the sequence submission
    corresponding to the overlapping clone, as we submit sequences with
    only a small overlap as described above.
    The following abbreviations are used to associate primary accession
    numbers given in the feature table with their source databases:
    Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information
    on the WormPEP database can be found at
    http://www.sanger.ac.uk/projects/c_elegans/wormpep This sequence
    has been finished according to sequence map criteria as follows.
```

An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. RPl-207J11 is from the library RPl-1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2

IMPORTANT: This sequence is the entire insert of clone RPl-207J11 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RPl-207J11 is at 1 in this sequence.

FEATURES

source

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

match: proteins: Tr:Q9Y2Q1 Tr:Q9U1I5 Sw:Q05481 Tr:Q02313 Sw:Q03923 Tr:Q43345 Tr:Q9Y2N8 Tr:Q62523 Sw:Q15122 Sw:Q9676 Sw:P28160 Tr:Q60792 Tr:Q9Y6R6 Sw:Q15928 Tr:Q62512 Tr:Q43693 Tr:Q9UBL4 Sw:Q75820 Tr:Q14586 Tr:Q95779 Sw:Q14628 Tr:Q95780 Sw:P49910 Tr:P70590 Sw:Q03924 Sw:P08045 Sw:P16415 Tr:Q61491 Tr:Q9ULS9 Sw:Q03936 Tr:Q88553 Tr:Q9QXT9 Tr:Q9UL37 Tr:Q9UG14 Tr:Q06054 Tr:Q14593 Tr:P97672 Tr:Q62510 Sw:P17097 /codon_start=1 /pseudo

/evidence=not_experimental

/note="AluY repeat: matches 1. .301 of consensus"

/note="AluY repeat: matches 1. .307 of consensus"

/note="AluY repeat: matches 1. .296 of consensus"

/note="L1MA2 repeat: matches 6250. .6306 of consensus"

/note="AluX repeat: matches 4. .273 of consensus"

/note="L1MA2 repeat: matches 6109. .6256 of consensus"

/note="AluSg repeat: matches 1. .312 of consensus"

/note="AluY repeat: matches 1. .303 of consensus"

/note="16 copies 2 mer at 93% conserved"

/note="AluY repeat: matches 1. .311 of consensus"

/note="L1P5 repeat: matches 6238. .6286 of consensus"

/note="AluSg repeat: matches 1. .307 of consensus"

/note="L1P5 repeat: matches 3193. .6238 of consensus"

/note="AluY repeat: matches 1. .311 of consensus"

/note="L1P5 repeat: matches 3036. .3193 of consensus"

/note="MER41A repeat: matches 1. .554 of consensus"

/note="MER57-internal repeat: matches 3. .1080 of consensus"

/note="AluSg repeat: matches 1. .303 of consensus"

/note="AluSg repeat: matches 1. .302 of consensus"

/note="MER57-internal repeat: matches 1228. .2371 of consensus"

/note="MER57-internal repeat: matches 2476. .2626 of consensus"

/note="HUERS-p3b repeat: matches 3446. .3510 of consensus"

/note="MER57-internal repeat: matches 3336. .3630 of consensus"

/note="MER51-internal repeat: matches 3587. .3767 of consensus"

/note="AluSg repeat: matches 1. .311 of consensus"

/note="MER51-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

/note="MER57-internal repeat: matches 3767. .4921 of consensus"

```
/note="MER41-internal repeat: matches 1449. .3944 of
consensus"
repeat_region 30673. .31219
/note="MER41A repeat: matches 1. .554 of consensus"
misc_feature complement(join(31775. .32111,32424. .32482))
/note="match: STS: Em:G61808"
repeat_region 32095. .32387
/note="AluSx repeat: matches 5. .297 of consensus"
repeat_region 33085. .33128
/note="MER4-internal repeat: matches 751. .794 of
consensus"
repeat_region 33491. .33698
/note="MER4-internal repeat: matches 465. .672 of
consensus"
repeat_region 33699. .33830
/note="AluSg repeat: matches 3. .138 of consensus"
repeat_region 33831. .34136
/note="AluSg repeat: matches 1. .303 of consensus"
repeat_region 34137. .34321
/note="AluSg repeat: matches 138. .309 of consensus"
repeat_region 34322. .34433
/note="MER4-internal repeat: matches 354. .465 of
consensus"

Query Match 0.3%; Score 77; DB 89; Length 79666;
Best Local Similarity 100.0%; Pred. No. 1.5e-30;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17364 tgagacggagtttcaactctgtgtgcccaggctggagtgcaatggcgtagctcagctcac 17423
|||||
Db 16601 TGAGACGGAGTTTCACTCTGTGTTGCCAGGCTGGAGTGCATGCGGTGATCTCAGCTCAC 16660
|||||

QY 17424 tgcaacctccgcctccc 17440
|||||
Db 16661 TGCAACCTCCGCCTCCC 16677

RESULT 32
AC053493/C
LOCUS AC053493 149496 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 12 clone RP11-230B21, WORKING DRAFT
SEQUENCE, 12 unordered pieces.
ACCESSION AC053493
VERSION AC053493.4 GI:7770025
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 149496)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
JOURNAL
AUTHORS Waterston,R.H.
REFERENCE 2 (bases 1 to 149496)
Waterston,R.H.
Direct Submission
TITLE Submitted (16-APR-2000) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On May 11, 2000 this sequence version replaced gi:7715653.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0230B21
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing method: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
```

```
Consensus quality: 143624 bases at least Q40
Consensus quality: 145128 bases at least Q30
Consensus quality: 146083 bases at least Q20
Insert size: 147000; agarose-fp
Insert size: 148396; sum-of-contigs
Quality coverage: 4.60 in Q20 bases; agarose-fp
Quality coverage: 4.61 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2365: contig of 2365 bp in length
* 2366 2465: gap of unknown length
* 2466 6704: contig of 4239 bp in length
* 6705 6804: gap of unknown length
* 6805 13071: contig of 6267 bp in length
* 13072 13171: gap of unknown length
* 13172 18348: contig of 5177 bp in length
* 18349 18448: gap of unknown length
* 18449 27648: contig of 9200 bp in length
* 27649 27748: gap of unknown length
* 27749 33840: contig of 6092 bp in length
* 33841 33940: gap of unknown length
* 33941 41084: contig of 7124 bp in length
* 41085 41164: gap of unknown length
* 41165 50648: contig of 9484 bp in length
* 50649 50748: gap of unknown length
* 50749 67548: contig of 16800 bp in length
* 67549 67648: gap of unknown length
* 67649 87565: contig of 19917 bp in length
* 87566 87666: gap of unknown length
* 87667 115777: contig of 28112 bp in length
* 115778 115877: gap of unknown length
* 115878 149496: contig of 33619 bp in length.

FEATURES
Location/Qualifiers
1. .149496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="12"
/clone="RP11-230B21"

BASE COUNT 46812 a 27161 c 27248 g 47164 t 1111 others
ORIGIN

Query Match 0.3%; Score 77; DB 72; Length 149496;
Best Local Similarity 100.0%; Pred. No. 1.6e-30;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4715 ttttgtatttttagtagacggggtttccaccatgttgccaggatagctcgatctct 4774
|||||
Db 72264 TTTTGTATTTTATAGTAGACGGGTTTCCACCATGTGSCCAGGATAGTCTCGATCTCT 72205
|||||

QY 4775 tgacctgtgactgcc 4791
|||||
Db 72204 TGACCTGTGATCTGCC 72188

RESULT 33
AC073283
LOCUS AC073283 175197 bp DNA HTG 07-AUG-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-761B3, WORKING DRAFT SEQUENCE,
21 unordered pieces.
ACCESSION AC073283
VERSION AC073283.5 GI:9719833
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```


REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 175197)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 175197)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 7, 2000 this sequence version replaced gi:9653223.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0761B03
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator; 100% of reads
Assembly: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165414 bases at least Q40
Consensus quality: 168975 bases at least Q30
Consensus quality: 170804 bases at least Q20
Insert size: 156000; agarose-fp
Insert size: 173197; sum-of-contigs
Quality coverage: 5.57 in Q20 bases; agarose-fp
Quality coverage: 5.05 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 1183: contig of 1183 bp in length
* 1184 1283: gap of unknown length
* 1284 4073: contig of 2790 bp in length
* 4074 4173: gap of unknown length
* 4174 7642: contig of 3469 bp in length
* 7643 7742: gap of unknown length
* 7743 10811: contig of 3069 bp in length
* 10812 10911: gap of unknown length
* 10912 13683: contig of 2772 bp in length
* 13684 13783: gap of unknown length
* 13784 17428: contig of 3645 bp in length
* 17429 17528: gap of unknown length
* 17529 21904: contig of 4376 bp in length
* 21905 22004: gap of unknown length
* 22005 27180: contig of 5176 bp in length
* 27181 27280: gap of unknown length
* 27281 30585: contig of 3305 bp in length
* 30586 30686: gap of unknown length
* 30686 36202: contig of 5517 bp in length
* 36203 36302: gap of unknown length
* 36303 45053: contig of 8751 bp in length
* 45054 45153: gap of unknown length
* 45154 54422: contig of 9269 bp in length
* 54423 54522: gap of unknown length
* 54523 62067: contig of 7545 bp in length
* 62068 62167: gap of unknown length
* 62168 71923: contig of 9756 bp in length
* 71924 72023: gap of unknown length
* 72024 81310: contig of 9287 bp in length
* 81311 81410: gap of unknown length
* 81411 89197: contig of 7787 bp in length
* 89198 89297: gap of unknown length

* 89298 103965: contig of 14668 bp in length
* 103966 104066: gap of unknown length
* 104066 115054: contig of 10989 bp in length
* 115055 115154: gap of unknown length
* 115155 128256: contig of 13102 bp in length
* 128257 128356: gap of unknown length
* 128357 142378: contig of 14022 bp in length
* 142379 142478: gap of unknown length
* 142479 175197: contig of 32719 bp in length.

FEATURES

source
1..175197
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-761B3"
misc_feature
1..1183
/note="assembly_name:Contig36"
misc_feature
1284..4073
/note="assembly_name:Contig40"
misc_feature
4174..7642
/note="assembly_name:Contig41"
misc_feature
7743..10811
/note="assembly_name:Contig42"
misc_feature
10912..13683
/note="assembly_name:Contig43"
misc_feature
13784..17428
/note="assembly_name:Contig44"
misc_feature
17529..21904
/note="assembly_name:Contig45"
misc_feature
22005..27180
/note="assembly_name:Contig46"
misc_feature
27281..30585
/note="assembly_name:Contig47"
misc_feature
30686..36202
/note="assembly_name:Contig48"
misc_feature
36303..45053
/note="assembly_name:Contig49"
misc_feature
45154..54422
/note="assembly_name:Contig50"
misc_feature
54523..62067
/note="assembly_name:Contig51"
misc_feature
62168..71923
/note="assembly_name:Contig52"
misc_feature
72024..81310
/note="assembly_name:Contig53"
misc_feature
81411..89197
/note="assembly_name:Contig54"
misc_feature
89298..103965
/note="assembly_name:Contig55"
misc_feature
104066..115054
/note="assembly_name:Contig56"
misc_feature
115155..128256
/note="assembly_name:Contig57"
misc_feature
128357..142378
/note="assembly_name:Contig58"
misc_feature
142479..175197
/note="assembly_name:Contig59"
BASE COUNT 46774 a 40217 c 38462 g 47738 t 2006 others
ORIGIN

Query Match 0.3%; Score 77; DB 74; Length 175197;
Best Local Similarity 100.0%; Pred. No. 1.7e-30;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4715 ttttttttttttagtagagacgggtttccacatgttggcaggatgtctcgtctct 4774
Db 32277 TTTTGTATTTTATTAGAGACGGGTTTCACCATGTTGCCAGGATGCTCTCTCT 32336
Qy 4775 tgacctgtgtatctgcc 4791
Db 32337 TGACCTTGTGATCTGCC 32353

RESULT 34

AC084754/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC084754 176626 bp DNA PRI 01-JAN-2001
Homo sapiens 12p BAC RP11-874G11 (Roswell Park Cancer Institute
Human BAC Library) complete sequence.

AC084754 GI:12000447

HTG.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 176626)

Alsbrooks, S. L., Amaratunga, H. C., Ali-osman, F. R., Allen, C.,

Benton, J., Blum, K., Blankenburg, K., Bonnin, D., Bouck, J.,

Bowles, S., Briveau, M., Brown, M., Bryant, N. P., Buhay, C.,

Burch, P., Burkett, C., Burrell, K. L., Byrd, N. C., Carron, T. F.,

Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R.,

Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C. D.,

Cox, C., Coyle, M. D., Dathorne, S. R., David, R., Davila, M. L., Davis, C.,

Davy-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O.,

Denn, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H.,

Dugan-Rocha, S., Durbin, K. J., Earnhart, C., Edgar, D., Edwards, C. C.,

Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D.,

Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J.,

Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W.,

Gundaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K.,

Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O.,

Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F.,

Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L. E.,

Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,

Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,

Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L. C.,

Lewis, L., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,

Louise, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R.,

Ma, J., Maheshwari, M., Mapua, P., Marondel, I., Martin, R.,

Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M. P.,

Meador, M., Mei, G., Merscher, S., Metzker, M., Miller, A., Miner, G.,

Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K. T., Morgan, M.,

Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N.,

Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwen, S.,

Ogih, M., Okuwonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,

Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L.,

Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M.,

Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shim, C.,

Shooshari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A.,

Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A.,

Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B.,

Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D.,

Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R.,

Washington, C., Watlington, S., Williams, G., Williamson, A.,

Wierzyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J.,

Zorrilla, S., Zucherlapati, R., and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 176626)

Worley, K. C.

Direct Submission

Submitted (15-NOV-2000) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 176626)

Worley, K. C.

Direct Submission

Submitted (01-JAN-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jan 1, 2001 this sequence version replaced gi:11995557.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
Contig length: 176626
Phrap values in estimate: 175830
Average error rate (BCM-Phrap estimate): 0.84818e-05
Fraction of Phrap values less than 40 : 1.0105784
Number of consensus changing edits: 25
Number of N's in consensus : 0

----- Consensus changing edits -----
Position Original+Context Edited+Context
13638 gatacctttg(n)ctataattca gatacctttg(t)ctataattca
31729 ttaagataaa(n)ggattatcct ttaagataaa(g)ggattatcct
31739 ctaagggaag(n)atttatttaag ctaagggaag(a)atttatttaag
51964 atatatatac(n)ctattatatac atatatatac(a)ctattatatac
60346 taccacacag(n)acagacacta taccacacag(t)acagacacta
73890 aagattttta(n)tcatttttaa aagattttta(t)tcatttttaa
74094 ttttactgtg(n)tgagatatgt ttttactgtg(t)tgagatatgt
93692 atgttcata(n)ggttatggt atgttcata(g)ggttatggt
93697 gtccgctttt(n)tcgtgtgccc gtccgctttt(c)tcgtgtgccc
125566 cctaaatgc(n)gagataaata cctaaatgc(t)gagataaata
131313 ttctctcttc(c)ttttttttt ttctctcttc(t)ttttttttt
145142 catctattg(n)cgcttttcta catctattg(t)cgcttttcta
153486 cgtctttcta(n)aactgagaaa cgtctttcta(a)aactgagaaa
155105 ccaattctaa(n)angnnngaaa ccaattctaa(g)agngngttaa
155107 aatcttaana(n)gngnggaagg aatcttaana(g)gngnggaagg
155109 tcttaanang(n)nggaaggaaa tcttaanang(g)nggaaggaaa
155110 cttaanang(n)nggaaggaaa cttaanang(g)nggaaggaaa
155111 ttaanangnn(n)ggaaggaaac ttaanangnn(g)ggaaggaaac
156603 tgtttattta(n)gcattgacct tgtttattta(t)gcattgacct
172601 cacttatagg(n)gggaattgaa cacttatagg(t)gggaattgaa
172670 tgttgtgggg(n)ngggggnnng tgttgtgggg(t)ngggggnnng
172671 ggttgtgggg(n)ggggggnnng ggttgtgggg(t)ggggggnnng
172677 gggnnngggg(n)ggggggaggg gggnnngggg(a)ggggggaggg
172678 ggtggggggg(n)ggggggaggg ggtggggggg(g)ggggggaggg


```
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17364 tgagacgaggttcaactctgttgccaggctgagtgcaatggcgtgatctcagctcac 17423
|||||
Db 29173 TGAGACGAGGTTTCACCTCTGTGTGCCAGGCTGGAGTGCATGGCGTGAATCTCAGCTCAC 29114
|||||
QY 17424 tgcaactccgcctccc 17440
|||||
Db 29113 TGCAACCTCGCCTCCC 29097
|||||

RESULT 36
AL137849
LOCUS
DEFINITION
AL137849 131684 bp DNA PRI 07-APR-2001
Human DNA sequence from clone RP11-507D14 on chromosome
9q21.2-22.1, complete sequence.
ACCESSION
AL137849
VERSION
AL137849.13 GI:13561124
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 131684)
Sehra,H.
Direct Submission
Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clones@requests.sanger.ac.uk
On Apr 8, 2001 this sequence version replaced gi:13396339.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr-9
RP11-507D14 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-507D14 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-507D14 is at 1 in this sequence.
The true left end of clone RP11-229M1 is at 131585 in this
sequence. The true right end of clone RP11-280P22 is at 65976 in
this sequence.
Location/Qualifiers
1..131684
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="9"
/map="q21.2-22.1"
/clone="RP11-507D14"
/clone_lib="RPCI-11.2"
1..69
/note="MER41B repeat: matches 567. .635 of consensus"

repeat_region
/note="match: STS: Em:G50245"
17856..18395
/note="119 copies 2 mer aa 71% conserved"
17477..17714
/note="AluX repeat: matches 4. .304 of consensus"
17182..17476
/note="L1MD3 repeat: matches 6537. .6595 of consensus"
16617..16675
/note="37 copies 2 mer tg 79% conserved"
16543..16616
/note="L1MC3 repeat: matches 6610. .6776 of consensus"
16367..16531
/note="MSTD repeat: matches 307. .394 of consensus"
16279..16366
/note="FLAM.A repeat: matches 4. .119 of consensus"
16164..16278
/note="AluSg repeat: matches 1. .303 of consensus"
14744..15046
/note="23 copies 2 mer aa 73% conserved"
14440..14485
/note="MSTB repeat: matches 1. .426 of consensus"
13997..14416
/note="L2 repeat: matches 1596. .1786 of consensus"
13551..13728
/note="MLT2FB repeat: matches 1. .392 of consensus"
13059..13482
/note="L2 repeat: matches 1920. .2147 of consensus"
12814..13041
/note="AluSg/x repeat: matches 135. .302 of consensus"
12325..12563
/note="L2 repeat: matches 2271. .2520 of consensus"
11975..12109
/note="AluX repeat: matches 1. .291 of consensus"
11597..11898
/note="AluSx repeat: matches 1. .293 of consensus"
11054..11346
/note="MSTB repeat: matches 1. .426 of consensus"
10088..10502
/note="MSTB-internal repeat: matches 1. .1651 of consensus"
8496..10087
/note="MSTB repeat: matches 373. .426 of consensus"
8442..8495
/note="AluSg repeat: matches 1. .280 of consensus"
8146..8425
/note="MSTB repeat: matches 1. .386 of consensus"
7770..8144
/note="L1M repeat: matches 4763. .5327 of consensus"
7196..7759
/note="L1 repeat: matches 3127. .3925 of consensus"
6413..7207
/note="AluSg/x repeat: matches 150. .310 of consensus"
6252..6412
/note="L1PBa repeat: matches 1189. .3130 of consensus"
5240..6251
/note="AluSg repeat: matches 5. .303 of consensus"
4936..5239
/note="L1PBa repeat: matches 762. .1189 of consensus"
4505..4935
/note="AluYb8 repeat: matches 1. .313 of consensus"
4191..4504
/note="L1PBa repeat: matches -279. .762 of consensus"
3142..4190
/note="L1PBa repeat: matches -940. .418 of consensus"
2610..3148
/note="L1PBa repeat: matches -1540. .1343 of consensus"
2412..2607
/note="68 copies 2 mer ga 81% conserved"
2364..2399
/note="MLT11 repeat: matches 58. .385 of consensus"
1149..1462
```



```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-280P22"
/clone_lib="RPC1-II.1"
1. 24780
/note="assembly_fragment:01918
fragment_chain:1
clone_end:SP6
vector_side:left"
24881. 33578
/note="assembly_fragment:00218
fragment_chain:1"
33679. 62574
/note="assembly_fragment:00251
fragment_chain:2"
62675. 84494
/note="assembly_fragment:00684
fragment_chain:2"
84595. 93785
/note="assembly_fragment:01499
fragment_chain:2"
93886. 104202
/note="assembly_fragment:02686
fragment_chain:2"
104303. 109696
/note="assembly_fragment:00215
fragment_chain:2"
109797. 115697
/note="assembly_fragment:00556
fragment_chain:2"
115798. 134601
/note="assembly_fragment:00926
fragment_chain:2
clone_end:T7
vector_side:right"
BASE COUNT 39424 a 28400 c 29106 g 36859 t 812 others
ORIGIN

Query Match 0.3%; Score 76; DB 81; Length 134601;
Best Local Similarity 100.0%; Pred. NO. 5.8e-30;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17555 caggctggtcgaactcctcagctgattccaccacccagctcccaaatgtt 17614
|||||
Db 127417 CAGGCTGGTCTCGAATCTCTGACCTCAGGTGATCCACCTCAGGCTCCCAAGTGT 127476

QY 17615 gggattacagggctga 17630
|||||
Db 127477 GGGATTACAGGCTGA 127492

RESULT 38
AC040997
LOCUS AC040997 149386 bp DNA HTG 12-MAY-2000
DEFINITION Homo sapiens chromosome 4 clone RP11-211G17 map 4, WORKING DRAFT
SEQUENCE, 19 unordered pieces.
ACCESSION AC040997
VERSION AC040997.2 GI:7770615
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 149386)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 4, clone RP11-211G17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 149386)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Balgwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehocsky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienda,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trifillio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 12, 2000 this sequence version replaced gi:7534199.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9525
Center clone name: 211.G.17
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 139979 bases at least Q40
Consensus quality: 143908 bases at least Q30
Consensus quality: 145966 bases at least Q20
Insert size: 151000; agarose-1p
Quality coverage: 4.3 in Q20 bases; agarose-1p
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1136: contig of 1136 bp in length
* 1137 1236: gap of 100 bp
* 1237 2859: contig of 1623 bp in length
* 2860 2959: gap of 100 bp
* 2960 4347: contig of 1388 bp in length
* 4348 4447: gap of 100 bp
* 4448 7318: contig of 2871 bp in length
* 7319 7418: gap of 100 bp
* 7419 9360: contig of 1942 bp in length
* 9361 9460: gap of 100 bp
* 9461 13630: contig of 4170 bp in length
* 13631 13730: gap of 100 bp
* 13731 16664: contig of 2934 bp in length
* 16665 16764: gap of 100 bp
* 16765 20769: contig of 4005 bp in length
* 20770 20869: gap of 100 bp
* 20870 24469: contig of 3600 bp in length
* 24470 24569: gap of 100 bp
* 24570 28677: contig of 4108 bp in length
* 28678 28777: gap of 100 bp

```

* 28778 33976: contig of 5199 bp in length
* 33977 34076: gap of 100 bp
* 34077 41823: contig of 7747 bp in length
* 41824 41923: gap of 100 bp
* 41924 50252: contig of 8329 bp in length
* 50253 50352: gap of 100 bp
* 50353 60535: contig of 10183 bp in length
* 60536 60635: gap of 100 bp
* 60636 72136: contig of 11501 bp in length
* 72137 72236: gap of 100 bp
* 72237 87733: contig of 15503 bp in length
* 87740 87839: gap of 100 bp
* 87840 102959: contig of 15120 bp in length
* 102960 103059: gap of 100 bp
* 103060 121780: contig of 18721 bp in length
* 121781 121880: gap of 100 bp
* 121881 149386: contig of 27506 bp in length.

FEATURES

source
1. .149386
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-211G17"
/clone_lib="RPCI-11 Human Male BAC"
1. .1136
/note="assembly_fragment"
1237. 2859
/note="assembly_fragment"
2960. 4347
/note="assembly_fragment"
4448. 7318
/note="assembly_fragment"
7419. 9360
/note="assembly_fragment"
9461. 13630
/note="assembly_fragment"
13731. 16664
/note="assembly_fragment"
clone_end:SP6
vector_side:left
16765. 20769
/note="assembly_fragment"
20870. 24469
/note="assembly_fragment"
24570. 28677
/note="assembly_fragment"
28778. 33976
/note="assembly_fragment"
34077. 41823
/note="assembly_fragment"
41924. 50252
/note="assembly_fragment"
50353. 60535
/note="assembly_fragment"
60636. 72136
/note="assembly_fragment"
72237. 87739
/note="assembly_fragment"
87840. 102959
/note="assembly_fragment"
103060. 121780
/note="assembly_fragment"
121881. 149386
/note="assembly_fragment"
BASE COUNT 47725 a 26017 c 25454 g 48389 t 1801 others
ORIGIN

Query Match 0.38; Score 76; DB 71; Length 149386;
Best Local Similarity 100.0%; Pred. No. 5.9e-30;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12635 aatttggatttttagtagacaggggttttccacgcttggtcaggtggtctcaaacctc 12694
|||||
Db 16398 AATTTGTATTTTAGTAGAGAGGGGTTTCCTCCAGGTGGTCAGGCTGCTCAAACTC 16457
|||||
Qy 12695 ctgacctcaggtgatc 12710
|||||
Db 16458 CTGACCTCAGGTGATC 16473
|||||
RESULT 39
AL391994 Homo sapiens chromosome 1 clone RP11-382D8, *** SEQUENCING IN
LOCUS PROGRESS ***, 9 unordered pieces.
DEFINITION ACCESSION
AL391994
VERSION AL391994.6 GI:13443435
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 165605)
Mclay,K.
Direct Submission
Submitted (19-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 24, 2001 this sequence version replaced gi:13373994.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA382D8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 162959 bases at least Q40
Consensus quality: 163679 bases at least Q30
Consensus quality: 164239 bases at least Q20
Insert size: 164805; sum-of-contigs
Insert size: 163443; 1.4% error; agarose-fp
Quality coverage: 6.92x in Q20 bases; sum-of-contigs Quality
coverage: 7.08x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 40047: contig of 40047 bp in length
* 40048 40147: gap of 100 bp
* 40148 54951: contig of 14804 bp in length
* 54952 55051: gap of 100 bp
* 55052 65031: contig of 9980 bp in length
* 65032 65131: gap of 100 bp
* 65132 74491: contig of 9360 bp in length
* 74492 74591: gap of 100 bp
* 74592 103412: contig of 28821 bp in length
* 103413 103512: gap of 100 bp
* 103513 108052: contig of 4540 bp in length
* 108053 108152: gap of 100 bp
* 108153 129520: contig of 21368 bp in length
* 129521 129620: gap of 100 bp
* 129621 133362: contig of 3742 bp in length
* 133363 133462: gap of 100 bp
* 133463 165605: contig of 32143 bp in length.
Location/Qualifiers
FEATURES


```
source
1..165605
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-382D8"
/clone_lib="RPC1-11.2"
1..40047
/note="assembly_fragment:01339
fragment_chain:1
clone_end:SP6
vector_side:left"
40148..54951
/note="assembly_fragment:03455
fragment_chain:1"
55052..65031
/note="assembly_fragment:02156
fragment_chain:1"
65132..74491
/note="assembly_fragment:03743
fragment_chain:1"
74592..103412
/note="assembly_fragment:01859
fragment_chain:2"
103513..108052
/note="assembly_fragment:02367
fragment_chain:2"
108153..129520
/note="assembly_fragment:00122
fragment_chain:3"
129621..133362
/note="assembly_fragment:00821
fragment_chain:3"
133463..165605
/note="assembly_fragment:00465
fragment_chain:3
clone_end:77
vector_side:right"
BASE COUNT 54164 a 33042 c 32203 g 45392 t 804 others
ORIGIN

Query Match 0.3%; Score 76; DB 81; Length 165605;
Best Local Similarity 100.0%; Pred. No. 6e-30;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17358 ttttttgacgaggttcactcttggccaggctgagtgcaatggcgatctca 17417
|||||
Db 150074 TTTTGTGACGGAGTTTCACTCTGTTGTCAGGCTGAGTCAATGGCGTGTCTCA 150133
|||||

QY 17418 gtcctagcaacctcc 17433
|||||
Db 150134 GCTCACTGCAACCTCC 150149

RESULT 40
AL357556/c
LOCUS AL357556 170195 bp DNA HTG 19-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-517124, *** SEQUENCING IN
PROGRESS ***, 10 unordered pieces.
ACCESSION AL357556
VERSION AL357556.12 GI:12329363
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 170195)
AUTHORS Burton,J.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
COMMENT On Jan 21, 2001 this sequence version replaced gi:12227373.

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA517124
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 3% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 166912 bases at least Q40
Consensus quality: 167970 bases at least Q30
Consensus quality: 168576 bases at least Q20
Insert size: 169295; sum-of-contigs
Quality coverage: 6.15x in Q20 bases; agarose-fp
coverage: 6.35x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
2050: contig of 2050 bp in length
1
2051 2150: gap of 100 bp
2151 13173: contig of 11023 bp in length
13174 13273: gap of 100 bp
13274 38221: contig of 24948 bp in length
38222 38321: gap of 100 bp
38322 78329: contig of 40008 bp in length
78330 78429: gap of 100 bp
78430 94546: contig of 16117 bp in length
94547 94646: gap of 100 bp
94647 102356: contig of 7710 bp in length
102357 102456: gap of 100 bp
102457 108098: contig of 5642 bp in length
108099 108198: gap of 100 bp
108199 125531: contig of 17333 bp in length
125532 125631: gap of 100 bp
125632 145739: contig of 20108 bp in length
145740 145839: gap of 100 bp
145840 170195: contig of 24356 bp in length.
Location/Qualifiers
1..170195
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-517124"
/clone_lib="RPC1-11.2"
1..2050
/note="assembly_fragment:01613
clone_end:SP6
vector_side:left"
2151..13173
/note="assembly_fragment:03000
fragment_chain:1"
13274..38221
/note="assembly_fragment:01701
fragment_chain:1"
38322..78329
/note="assembly_fragment:00873"
78430..94546
/note="assembly_fragment:03071"
94647..102356
/note="assembly_fragment:01722
fragment_chain:2"
102457..108098
/note="assembly_fragment:01472

FEATURES
source
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
```

```
fragment_chain:2"
108199..125531
/note="assembly_fragment:00163
fragment_chain:2"
125632..145739
/note="assembly_fragment:01151
fragment_chain:2"
145840..170195
/note="assembly_fragment:02163
fragment_chain:2
clone_end:77
vector_side:right"
BASE COUNT 49251 a 37221 c 36147 g 46669 t 907 others
ORIGIN

Query Match 0.3%; Score 76; DB 80; Length 170195;
Best Local Similarity 100.0%; Pred. No. 6e-30;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17358 ttttttgagcggatttcactctgttgcacgctggagtgcaatggcgtgatctca 17417
|||||
Db 162512 tttttttgagcggagtgatttcactctgttgcacgctggagtgcaatggcgtgatctca 162453
|||||

QY 17418 gctcactgcaacctcc 17433
|||||
Db 162452 gctcactgcaacctcc 162437
|||||

RESULT 41
LOCUS HS281H8/c
DEFINITION HS281H8 110414 bp DNA PRI 23-NOV-1999
Contains up to four novel genes, one with similarity to KIAA0323
and worm C30F12.1 and another with Ubiquitin-like protein gene SMT3
(the latter in an intron of a novel gene). Contains ESTs, STSS,
GSSs, a putative CpG island and genomic marker D6S1553, complete
sequence.
ACCESSION AL031133
VERSION AL031133.1 GI:3676189
KEYWORDS HTG; C30F12.1; CpG island; D6S1553; KIAA0323; SMT3; Ubiquitin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 110414)
Mashregni-Mohammadi,M.
Direct Submission
Submitted (09-NOV-1998) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Sep 30, 1998 this sequence version replaced gi:3646060.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence is the entire insert of clone 281H8. This sequence
has been finished according to sequence map criteria as follows. An
attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed by the Sanger Centre Chromosome 6
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
281H8 is from the library RPC11 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
Location/Qualifiers

source
1..110414
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q25.1-25.3"
/clone="RPI-281H8"
/clone_lib="RPC1-1"
1..62
repeat_region
/note="AluS repeat: matches 1..62 of consensus"
296..598
repeat_region
/note="AluSg repeat: matches 1..294 of consensus"
1982..2036
repeat_region
/note="MLTIB repeat: matches 336..390 of consensus"
2041..2156
repeat_region
/note="AluJo repeat: matches 1..116 of consensus"
2157..2463
repeat_region
/note="AluY repeat: matches 1..309 of consensus"
2494..3534
prim_transcript
/note="match: ESTs T61230 AA811212 AI085416 T61116
AA089548 AA282430 AI093951 AI097335 AA570035 AA095499
N56553 AA443227 W07118 AA493141 AI148021 AA465621 AI142307
AA578894 AA327805 AA096262 AA133846 AA843407 AI148678
T48328 AI188279 AI201880 AA262594 T48329 AA255682
AA443189"
complement(3150..3521)
/note="match: STS G41905"
3538..3735
repeat_region
/note="AluJo repeat: matches 105..305 of consensus"
3738..3978
repeat_region
/note="MLTIC repeat: matches 173..400 of consensus"
3979..4290
repeat_region
/note="AluSg repeat: matches 3..299 of consensus"
4291..4454
repeat_region
/note="MLTIC repeat: matches 6..173 of consensus"
4828..5037
repeat_region
/note="MIR repeat: matches 19..233 of consensus"
5184..5461
repeat_region
/note="AluJo repeat: matches 5..303 of consensus"
5480..5757
repeat_region
/note="L1MB3 repeat: matches 5290..5592 of consensus"
6256..6532
repeat_region
/note="AluSg repeat: matches 1..294 of consensus"
7396..7697
repeat_region
/note="AluSg repeat: matches 1..293 of consensus"
7799..7956
repeat_region
/note="MIR repeat: matches 19..186 of consensus"
8015..8306
repeat_region
/note="AluSg repeat: matches 1..309 of consensus"
8451..8745
repeat_region
/note="AluJo repeat: matches 1..299 of consensus"
8877..8949
repeat_region
/note="MIR repeat: matches 31..104 of consensus"
complement(9239..9391)
/note="match: STS G02971"
9503..9582
repeat_region
/note="L1M2 repeat: matches -493..-427 of consensus"
10792..10933
repeat_region
/note="L1MB3 repeat: matches 5687..5828 of consensus"
10935..11153
repeat_region
/note="AluJo repeat: matches 1..216 of consensus"
11158..11474
repeat_region
/note="L1MB3 repeat: matches 5860..6185 of consensus"
join(<11655..11734,20172..>20399)
/gene="dJ281H8.1"
/note="match: ESTs R16664 T82134
partially supported by FGENES and GENSCAN"
/evidence="not experimental"
/product="dJ281H8.1 (PUTATIVE novel protein)"
join(11655..11734,20172..>20399)
/gene="dJ281H8.1"
11878..11988
repeat_region
/note="MIR repeat: matches 32..152 of consensus"
12976..13062
repeat_region
```

```
/note="MIR repeat: matches 65. .167 of consensus"
13112. .13408
/Note="AluX repeat: matches 1. .298 of consensus"
13589. .13906 repeat: matches -416. -.78 of consensus"
/Note="L1M3 repeat: matches 3. .311 of consensus"
14021. .14300
/Note="AluSp repeat: matches 3. .311 of consensus"
14304. .14507
/Note="L1PA15 repeat: matches 5774. .5985 of consensus"
14710. .14931
/Note="L2 repeat: matches 1763. .2029 of consensus"
14952. .15082
/Note="FLAM_A repeat: matches 4. .132 of consensus"
15219. .15388
/Note="L2 repeat: matches 2382. .2551 of consensus"
15389. .15705
/Note="AluJ repeat: matches 3. .312 of consensus"
15706. .15934
/Note="L2 repeat: matches 2124. .2383 of consensus"
16167. .16599
/Note="Charlie1b repeat: matches 74. .511 of consensus"
16600. .16837
/Note="AluSg repeat: matches 1. .299 of consensus"
16898. .16981
/Note="Charlie1b repeat: matches 1. .74 of consensus"
17163. .17513
/Note="L2 repeat: matches 1206. .1570 of consensus"
17826. .17918
/Note="L2 repeat: matches 2623. .2710 of consensus"
17972. .18259
/Note="AluJ repeat: matches 1. .288 of consensus"
18269. .18560
/Note="AluSg repeat: matches 1. .293 of consensus"
19058. .19349
/Note="AluSp repeat: matches 1. .295 of consensus"
19529. .19847
/Note="MLT1B repeat: matches 7. .376 of consensus"
complement(19861. .20234)
/Note="match: GSS AQ199235"
20624. .20842
/Note="MIR repeat: matches 14. .261 of consensus"
23033. .23133
/Note="MLT1D repeat: matches 249. .359 of consensus"
23854. .24062
/Note="WER3 repeat: matches 1. .209 of consensus"
24066. .24114
/Note="MIR repeat: matches 59. .110 of consensus"
24228. .24407
/Note="AluSc repeat: matches 124. .303 of consensus"
24409. .24705
/Note="AluX repeat: matches 3. .301 of consensus"
24744. .24807
/Note="MIR repeat: matches 56. .121 of consensus"
25312. .25609
/Note="AluX repeat: matches 1. .298 of consensus"
26509. .26813
/Note="AluSp repeat: matches 5. .310 of consensus"
26754. .27134
/Note="match: EST AA731910"
27097. .27216
/Note="MIR repeat: matches 116. .252 of consensus"
27232. .27530
/Note="AluJ repeat: matches 1. .296 of consensus"
27880. .28189
/Note="AluSg repeat: matches 1. .310 of consensus"
29019. .29312
/Note="AluY repeat: matches 1. .293 of consensus"
29416. .29709
/Note="AluX repeat: matches 5. .298 of consensus"
29870. .30043
/Note="L2 repeat: matches 2251. .2415 of consensus"
30073. .30348
/Note="L2 repeat: matches 2428. .2710 of consensus"
```

```
repeat_region 30531. .31085
/Note="L2 repeat: matches 1449. .2046 of consensus"
31086. .31356
/Note="AluSg repeat: matches 1. .271 of consensus"
31357. .31424
/Note="L2 repeat: matches 2046. .2109 of consensus"
31718. .32264
/Note="MLT2D repeat: matches 1. .552 of consensus"
32383. .32693
/Note="AluX repeat: matches 1. .311 of consensus"
Join(32815. .32954,37885. .38119,42063. .42169)
/Note="dJ281H8.2"

Query Match 0.3%; Score 75; DB 92; Length 110414;
Best Local Similarity 100.0%; Pred. No. 2,1e-29;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12635 aatttgatttttagtagacggggtttctccacgttggtcaggctgtctcaactc 12694
|||||
Db 26638 AATTGTGATTTTAGTAGAGACGGGTTTCTCCAGTTGGTCAGGCTGCTCAAACTC 26579

QY 12695 ctgacctcagggtgat 12709
|||||
Db 26578 CTGACCTCAGGTGAT 26564

RESULT 42
HS28F12 120630 bp DNA PRI 18-MAR-2001
LOCUS Human DNA sequence from clone RP1-28F12 on chromosome 20q11.22-12
DEFINITION Contains part of the KIAA0823 gene, ESTs, STSs and GSSs, complete
sequence.
ACCESSION AL031657
VERSION AL031657.5 GI:115444571
KEYWORDS HTG; KIAA0823.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 120630)
AUTHORS Skuce,C.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Dec 4, 2000 this sequence version replaced gi:9795179.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP1-28F12 is from the library RPCI-1 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP1-28F12 The true left
end of clone RP5-1076E17 is at 58615 in this sequence. The true
right end of clone RP11-12201 is at 2837 in this sequence. This
sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
```

one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

FEATURES

source

1..120630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="q11.22-12"
/clone="RP1-28F12"
/clone_lib="RPC1-1"

repeat_region

/note="AluY repeat: matches 2. .62 of consensus"

misc_feature

785. .1139

misc_feature

/note="match: GSS: Em:AQ770843"

repeat_region

complement(817. .1337)

misc_feature

/note="match: GSS: Em:AQ786836"

repeat_region

834. .1116

misc_feature

/note="AluSg repeat: matches 1. .282 of consensus"

repeat_region

complement(2169. .2835)

misc_feature

/note="match: GSS: Em:AQ342910"

repeat_region

2446. .2705

repeat_region

/note="AluSx repeat: matches 53. .311 of consensus"

repeat_region

2726. .3021

repeat_region

/note="L1MC/D repeat: matches 5495. .5796 of consensus"

repeat_region

3041. .3166

repeat_region

/note="AluSg/x repeat: matches 182. .308 of consensus"

repeat_region

3241. .3532

repeat_region

/note="AluSx repeat: matches 1. .311 of consensus"

misc_feature

complement(3278. .3626)

misc_feature

/note="match: GSS: Em:AQ493285"

misc_feature

complement(3281. .3629)

misc_feature

/note="match: GSS: Em:AQ493295"

repeat_region

complement(3367. .3625)

repeat_region

/note="match: GSS: Em:B63323"

misc_feature

3573. .3716

misc_feature

/note="L1MC5 repeat: matches 7246. .7383 of consensus"

repeat_region

3632. .4156

repeat_region

/note="match: GSS: Em:AQ610322"

repeat_region

3907. .4124

repeat_region

/note="L1R33 repeat: matches 41. .256 of consensus"

repeat_region

4160. .4308

repeat_region

/note="L1MC5 repeat: matches 7375. .7522 of consensus"

repeat_region

4309. .4643

repeat_region

/note="AluSx repeat: matches 7. .312 of consensus"

repeat_region

/note="MLTIB repeat: matches 97. .390 of consensus"

repeat_region

11592. .11624

repeat_region

/note="MLTIF repeat: matches 244. .276 of consensus"

repeat_region

11680. .11745

repeat_region

/note="MLTIF repeat: matches 211. .276 of consensus"

repeat_region

11759. .11811

repeat_region

/note="MLTIF repeat: matches 238. .288 of consensus"

repeat_region

11993. .12118

repeat_region

/note="MLTIF repeat: matches 374. .529 of consensus"

repeat_region

12435. .12587

repeat_region

/note="MER46C repeat: matches 1. .149 of consensus"

repeat_region

12588. .12890

repeat_region

/note="AluSx repeat: matches 1. .305 of consensus"

repeat_region

12891. .12959

repeat_region

/note="MER46C repeat: matches 149. .215 of consensus"

repeat_region

12960. .13132

repeat_region

/note="MERSA repeat: matches 1. .189 of consensus"

repeat_region

13133. .13230

repeat_region

/note="MER46C repeat: matches 215. .330 of consensus"

repeat_region

13309. .13601

repeat_region

/note="AluSp repeat: matches 1. .294 of consensus"

repeat_region

13660. .13821

repeat_region

/note="MIR repeat: matches 90. .262 of consensus"

repeat_region

13841. .13874

repeat_region

/note="LTR33 repeat: matches 235. .268 of consensus"

repeat_region

14160. .14172

repeat_region

/note="L1ME3 repeat: matches 5914. .5926 of consensus"

repeat_region

14173. .14492

repeat_region

/note="AluSg repeat: matches 1. .313 of consensus"

repeat_region

14493. .15022

repeat_region

/note="L1ME3 repeat: matches 5384. .5914 of consensus"

repeat_region

15082. .15170

repeat_region

/note="TIGER2 repeat: matches 1. .77 of consensus"

repeat_region

15719. .15743

repeat_region

/note="L2 repeat: matches 2685. .2709 of consensus"

repeat_region

15744. .15884

repeat_region

/note="AluJo repeat: matches 2. .148 of consensus"

repeat_region

15885. .16178

repeat_region

/note="AluSx repeat: matches 1. .294 of consensus"

repeat_region

16179. .16360

repeat_region

/note="AluJo repeat: matches 148. .309 of consensus"

repeat_region

16361. .16459

repeat_region

/note="L2 repeat: matches 2584. .2685 of consensus"

repeat_region

16460. .16730

repeat_region

/note="MLTIB repeat: matches 115. .390 of consensus"

repeat_region

16731. .17043

repeat_region

/note="AluSx repeat: matches 1. .312 of consensus"

repeat_region

17044. .17159

repeat_region

/note="MLTIB repeat: matches 1. .115 of consensus"

repeat_region

17160. .17378

repeat_region

/note="L2 repeat: matches 2376. .2584 of consensus"

repeat_region

17594. .17605

repeat_region

/note="MERSA repeat: matches 150. .161 of consensus"

repeat_region

17606. .17728

repeat_region

/note="L1MC4 repeat: matches 7854. .7977 of consensus"

repeat_region

17777. .17923

repeat_region

/note="AluJo repeat: matches 1. .141 of consensus"

repeat_region

17924. .18228

repeat_region

/note="AluY repeat: matches 1. .305 of consensus"

repeat_region

18229. .18372

repeat_region

/note="AluJo repeat: matches 141. .302 of consensus"

repeat_region

18512. .19040

repeat_region

/note="L1MC4 repeat: matches 7071. .7608 of consensus"

repeat_region

19081. .19148

repeat_region

/note="34 copies 2 mer ct 66% conserved"

repeat_region

19152. .19462

repeat_region

/note="AluSx repeat: matches 1. .311 of consensus"

repeat_region

19463. .19495

repeat_region

/note="MERSA repeat: matches 44. .79 of consensus"

repeat_region

19731. .20171

repeat_region

/note="MSTB repeat: matches 2. .422 of consensus"

repeat_region

20325. .20429

repeat_region

/note="MERSB repeat: matches 1. .106 of consensus"

repeat_region

11233. .11520

```
repeat_region 20469..20772
/note="AluY repeat: matches 1. .304 of consensus"
repeat_region 21337..21646
/note="AluJo repeat: matches 1. .306 of consensus"
repeat_region 21657..21946
/note="AluX repeat: matches 1. .303 of consensus"
repeat_region 22079..22372
/note="AluY repeat: matches 1. .294 of consensus"
repeat_region 22697..23131

Query Match 0.3%; Score 75; DB 92; Length 120630;
Best Local Similarity 100.0%; Pred. No. 2.1e-29;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17559 ctggtctgaactcctgacctcaggatgccaccaccctcagcctcccaaaagtgtggaa 17618
|||||
Db 90971 CTGGTCTGAACCTCCAGCTCAGGTGATCCACCCTCAGCTCCCAAGGTGTGGGA 91030
|||||

QY 17619 ttacaggcgtgagcc 17633
|||||
Db 91031 TTACAGCGGTGAGCC 91045
|||||

RESULT 43
AC008556
LOCUS AC008556 142515 bp DNA PRI 28-JUL-2000
DEFINITION Homo sapiens chromosome 19 clone CTC-526N19, complete sequence.
ACCESSION AC008556
VERSION AC008556.5 GI:9558574
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 142515)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 142515)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 142515)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 28, 2000 this sequence version replaced gi:7711297.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.6% of Sequence;
Estimated Total Number of Errors is 0.6.
STS Content:
SHGC-32254 G27463
WI-13386 G21725.
FEATURES
source
1..142515
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTC-526N19"
BASE COUNT 35038 a 33968 c 33523 g 38986 t
ORIGIN

Query Match 0.3%; Score 75; DB 86; Length 142515;
Best Local Similarity 100.0%; Pred. No. 2.1e-29;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8486 tgggtgaaccctgctgtactaaaaatacaaaaattagctgggtgtggcgcatgcct 8545
```

```
|||||
Db 110391 TGGTGAACCCCTGCTGTACTAAAAATACAAAAATTAGCTGGTGGTGGCGCATGCCT 110450
|||||
QY 8546 gtaatcccaagtact 8560
|||||
Db 110451 GTAATCCCAAGTACT 110465
|||||

RESULT 44
AC004534
LOCUS AC004534 155952 bp DNA PRI 03-FEB-2000
DEFINITION Homo sapiens PAC clone RP4-545C24 from 7q21-q22, complete sequence.
ACCESSION AC004534
VERSION AC004534.1 GI:3041858
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155952)
AUTHORS Dauphin, S. and Biewald, T.
TITLE The sequence of Homo sapiens PAC clone RP4-545C24
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155952)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (09-APR-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
3 (bases 1 to 155952)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
-----
Center project name: H_DJ0545C24
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
This clone was provided for sequencing by Dr. Stephen Scherer,
Department of Genetics, The Hospital for Sick Children, Toronto,
Ontario, Canada, with support from the Canadian Genome Analysis and
Technology Program; and Dr. John D. McPherson, Department of
Genetics, Washington University, St. Louis MO. For additional
information about the map position of this sequence, see
http://www.genet.sickkids.on.ca/chromosome7 and
http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-4, prepared by
Pieter de Jong and coworkers at Roswell Park Cancer Institute,
using the method described by Ioannou et al., Nature Genetics
6:84-9 (1994). The library is from one male donor. For further
details, see http://bacpac.med.buffalo.edu/
The clone is available from Genome Systems, Inc.
(http://www.genomesystems.com).
VECTOR: pCYPAC2
```



```

QY 10409 cctctgctccaccagg 10423
|||||
Db 140907 CCTCTGCTCCACAGG 140921

RESULT 45
HS26H23          91835 bp      DNA          22-NOV-1999
LOCUS              Human DNA sequence from PAC 26H23, BRCA2 gene region chromosome
DEFINITION        13q12-13 contains ESTs, CpG island.
ACCESSION        Z84467.285990 Z85991 Z85992 Z85993
VERSION          Z84467.1  G1:2104578
KEYWORDS         13q12-13; CpG island.
SOURCE           human.
ORGANISM          Homo sapiens
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE         1 (bases 1 to 91835)
AUTHORS           Williamson,H.
TITLE             Direct Submission
JOURNAL           Submitted (06-MAY-1997) Sanger Centre, Hinxton, Cambridgeshire,
                  CB10 1RQ, UK. E-mail enquiries: humquery@sanger.ac.uk
                  requests: clonerequest@sanger.ac.uk
                  On May 17, 1997 this sequence version replaced gi:1806009.
COMMENT           IMPORTANT:
                  This sequence is not the entire insert of clone 26H23. It may be
                  shorter because we only sequence overlapping sections once, or
                  longer because we arrange for a small overlap between neighbouring
                  submissions.
                  This sequence has been finished according to sequence map criteria
                  as follows. An attempt is made to resolve all sequencing problems,
                  such as compressions and repeats, but not necessarily within known
                  annotated human repeat sequence elements (e.g. Alu). Where the
                  sequence is ambiguous, there is an annotation using the 'unsure'
                  feature key.
                  The true left end of clone 26H23 is at 1 in this sequence. The true
                  right end of clone 130N4 is at 34585.
                  The true left end of clone 267P19 is at 91732.
                  26H23 is from the human PAC library described in Ioannou A.P. et al
                  Nature Genet 6, 84-89.
FEATURES          Location/Qualifiers
source            1..91835
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="13"
                  /map="13q12-13"
                  /clone_lib="RPC1-1"
                  /clone="XX-26H23"
                  1..83
repeat_region    /note="AluSc repeat: matches 84..1 of consensus;
                  incomplete repeat"
repeat_region    94..396
repeat_region    /note="AluSx repeat: matches 302..1 of consensus"
                  451..710
repeat_region    /note="L1 repeat: matches 4886..5146 of consensus"
                  708..815
repeat_region    /note="L1MB4 repeat: matches 2..91 of consensus"
                  817..1109
repeat_region    /note="AluSg repeat: matches 299..5 of consensus"
                  1113..1165
repeat_region    /note="L1MA4 repeat: matches 87..138 of consensus"
                  1195..1249
repeat_region    /note="AluJ repeat: matches 1..55 of consensus; incomplete
                  repeat"
                  1252..1551
repeat_region    /note="AluSg repeat: matches 3..303 of consensus"
                  1593..1797
repeat_region    /note="MSTD repeat: matches 185..391 of consensus"
                  1669..1796
repeat_region    /note="MSTB repeat: matches 289..422 of consensus"
                  3346..3416
repeat_region    /note="L1MB4 repeat: matches 859..929 of consensus"
                  3418..3552

```

```

/note="L1 repeat: matches 4982..5119 of consensus"
3558..3838
/note="AluSg repeat: matches 1..282 of consensus;
incomplete repeat"
4035..4117
/note="L1ME3 repeat: matches 5..76 of consensus"
4120..4429
/note="AluJb repeat: matches 3..302 of consensus"
4561..5055
/note="L1MB8 repeat: matches 69..572 of consensus"
5056..5186
/note="FLAM_C repeat: matches 1..132 of consensus"
5188..5500
/note="L1MB3 repeat: matches 567..923 of consensus"
6890..7101
/note="MIR repeat: matches 15..228 of consensus"
7430..7715
/note="AluJb repeat: matches 1..292 of consensus"
7718..8028
/note="AluSx repeat: matches 1..302 of consensus"
8460..8763
/note="AluSx repeat: matches 302..1 of consensus"
8979..9338
/note="MLT1A1 repeat: matches 1..365 of consensus"
9635..9799
/note="L1MA5A repeat: matches 636..472 of consensus"
9670..9814
/note="L1PA15 repeat: matches 600..458 of consensus"
11277..11857
/note="L1PA15 repeat: matches 311..902 of consensus"
12059..12316
/note="RHELC repeat: matches 371..121 of consensus"
12476..12670
/note="AluJo repeat: matches 301..107 of consensus;
incomplete repeat"
12673..12959
/note="AluSg repeat: matches 286..1 of consensus"
12968..13106
/note="FLAM_C repeat: matches 133..2 of consensus"
13148..13425
/note="AluSg repeat: matches 1..300 of consensus"
13562..14154
/note="L1ME1 repeat: matches 610..3 of consensus"
14009..14231
/note="L1 repeat: matches 5390..5167 of consensus"
14258..14414
/note="L1MA4 repeat: matches 742..905 of consensus"
14483..14608
/note="FLAM_A repeat: matches 1..126 of consensus"
14620..14736
/note="L1MA4A repeat: matches 930..1042 of consensus"
15523..15816
/note="AluSc repeat: matches 294..1 of consensus"
15867..16161
/note="AluSx repeat: matches 1..293 of consensus"
16162..16364
/note="L1 repeat: matches 4313..4112 of consensus"
16668..16953
/note="AluY repeat: matches 1..295 of consensus"
17125..17538
/note="match: multiple ESTs; low % ID; match: T59663
R48957 H24241 CL4861 F03338; match: W88836 N22360 H70851
F02679 N35317; match: R10942 N73317 N20948 R38018 R42572;
match: R62472 R43248 R41198 R38482 R42487; match: R05993
N71787 T64627 F03312"
18303..18654
/note="MLT1A1 repeat: matches 1..357 of consensus"
18784..19084
/note="AluSx repeat: matches 2..302 of consensus"
20018..20312
/note="AluSg repeat: matches 298..2 of consensus"
21304..21583
/note="AluSx repeat: matches 289..1 of consensus"

```



```

HS69M21/c
LOCUS       HS69M21      125937 bp      DNA      PRI      23-NOV-1999
DEFINITION   Human DNA sequence from clone 69M21 on chromosome 1p36.21-36.23
              Contains ESTs and CDSs, complete sequence.
ACCESSION   AL0311735
VERSION     AL0311735.9  GI:5102570
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 125937)
AUTHORS     Bagguley,C.
TITLE       Direct Submission
JOURNAL     Submitted (21-APR-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequests@sanger.ac.uk
COMMENT     On Jun 19, 1999 this sequence version replaced gi:4582139.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre Chromosome 1
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            69M21 is from the library RPCI1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/VECTOR: pCYPAC2
            This sequence is the entire insert of clone 69M21. The true right
            end of clone 330012 is at 31830 in this sequence.

FEATURES             Location/Qualifiers
     source           1..125937
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p36.21-36.23"
                     /clone="RPI-69M21"
                     /clone_lib="RPCI-1"
     repeat_region    1..217
                     /note="AluSx repeat: matches 21..237 of consensus"
     repeat_region    616..914
                     /note="AluSx repeat: matches 1..298 of consensus"
     repeat_region    2045..2215
                     /note="MER69 repeat: matches 37..198 of consensus"
     repeat_region    2223..2264
                     /note="MER69 repeat: matches 718..763 of consensus"
     repeat_region    2265..2556
                     /note="MLTID repeat: matches 169..505 of consensus"
     repeat_region    2557..2855
                     /note="AluSx repeat: matches 1..304 of consensus"
     repeat_region    2856..3036
                     /note="MLTID repeat: matches 1..169 of consensus"
     repeat_region    3037..3162
                     /note="MER69 repeat: matches 763..867 of consensus"
     repeat_region    3206..3380
                     /note="AluB repeat: matches 131..294 of consensus"
     repeat_region    3391..3556
                     /note="MER69 repeat: matches 2137..2318 of consensus"
     repeat_region    3640..3923
                     /note="AluB repeat: matches 5..288 of consensus"
     repeat_region    3956..4085
                     /note="I2 repeat: matches 2358..2488 of consensus"
     repeat_region    4086..4372
                     /note="AluSg repeat: matches 1..299 of consensus"
     repeat_region    4373..4578
                     /note="I2 repeat: matches 2488..2691 of consensus"
     repeat_region    4662..4714
                     /note="MIR repeat: matches 48..100 of consensus"
     repeat_region    5433..5469
                     /note="MER91 repeat: matches 1..38 of consensus"
     repeat_region    5718..5805
                     /note="MIR repeat: matches 61..151 of consensus"
     repeat_region    6096..6124
                     /note="MIR repeat: matches 116..144 of consensus"
     repeat_region    6243..6272
                     /note="MIR repeat: matches 109..138 of consensus"
     repeat_region    6687..6817
                     /note="MIR repeat: matches 122..247 of consensus"
     repeat_region    7526..7653
                     /note="MIR repeat: matches 54..170 of consensus"
     repeat_region    7761..7818
                     /note="MER74A repeat: matches 476..531 of consensus"
     repeat_region    7819..8126
                     /note="AluSg repeat: matches 1..308 of consensus"
     repeat_region    8127..8585
                     /note="MER74A repeat: matches 3..476 of consensus"
     repeat_region    8589..8639
                     /note="MIR repeat: matches 63..113 of consensus"
     variation        9150..9158
                     /note="clone 330012; gttttttt in this entry; insertion"
                     /replace="gt"
     repeat_region    9151..9450
                     /note="AluSx repeat: matches 1..299 of consensus"
     repeat_region    9533..9586
                     /note="I27 copies 2 mer gt 87% conserved"
     repeat_region    9825..9925
                     /note="MIR repeat: matches 35..144 of consensus"
     repeat_region    10001..10183
                     /note="MER5B repeat: matches 2..187 of consensus"
     repeat_region    10574..10637
                     /note="MIR repeat: matches 83..144 of consensus"
     repeat_region    10917..11075
                     /note="MIR repeat: matches 8..172 of consensus"
     repeat_region    11158..11340
                     /note="MIR repeat: matches 63..262 of consensus"
     repeat_region    11827..11941
                     /note="MIR repeat: matches 47..127 of consensus"
     repeat_region    11942..12244
                     /note="AluB repeat: matches 1..296 of consensus"
     repeat_region    12245..12322
                     /note="MIR repeat: matches 127..214 of consensus"
     repeat_region    14029..14111
                     /note="MIR repeat: matches 63..146 of consensus"
     repeat_region    14140..14433
                     /note="MLT2FB repeat: matches 2..264 of consensus"
     repeat_region    14434..14797
                     /note="THE1B repeat: matches 1..364 of consensus"
     repeat_region    14798..14949
                     /note="MLT2FB repeat: matches 263..414 of consensus"
     repeat_region    16256..16405
                     /note="MIR repeat: matches 32..173 of consensus"
     repeat_region    16506..16582
                     /note="MIR repeat: matches 185..262 of consensus"
     repeat_region    16603..16675
                     /note="MLT1J repeat: matches 61..135 of consensus"
     misc_feature     complement(<17679..>18271)
     repeat_region    17919..18552
                     /note="match: GSS AQ039685 clone 2325111"
     repeat_region    18553..19270
                     /note="ITRI2 repeat: matches 1..1481 of consensus"
     repeat_region    18562..18564
                     /note="PTR5 repeat: matches 743..1481 of consensus"
     variation        18562..18564
                     /note="clone 330012; tcy in this entry; substitution"
                     /replace="ttg"
     repeat_region    19278..19433
                     /note="MLT1J repeat: matches 211..371 of consensus"
     repeat_region    19875..20175

```

repeat_region /note="AluSg1 repeat: matches 1. .306 of consensus"
20188. .20226
variation /note="L1MB6 repeat: matches 6134. .6172 of consensus"
20226. .20227
/note="clone 330012; ct in this entry; deletion"
/replace="ctt"
repeat_region /note="AluSx repeat: matches 1. .299 of consensus"
20227. .20522
repeat_region /note="AluSg repeat: matches 5766. .6134 of consensus"
20523. .20892
repeat_region /note="AluJo repeat: matches 1. .301 of consensus"
20899. .21186
repeat_region /note="L2 repeat: matches 2137. .2418 of consensus"
21192. .21465
repeat_region /note="L1MC5 repeat: matches 7260. .7791 of consensus"
21542. .22135
repeat_region /note="L1MC5 repeat: matches 7260. .7791 of consensus"
22136. .22294
repeat_region /note="AluSc repeat: matches 150. .308 of consensus"
22344. .22421
repeat_region /note="L1MC5 repeat: matches 7163. .7239 of consensus"
22427. .22627
repeat_region /note="MER20 repeat: matches 1. .218 of consensus"
22853. .22987
repeat_region /note="AluJo repeat: matches 1. .135 of consensus"
22988. .23284
repeat_region /note="AluSg repeat: matches 1. .295 of consensus"
23285. .23428
repeat_region /note="AluJo repeat: matches 134. .280 of consensus"
23456. .23596
repeat_region /note="L1MD1 repeat: matches 5685. .5825 of consensus"
23662. .23901
repeat_region /note="L2 repeat: matches 2445. .2669 of consensus"
23905. .23940
repeat_region /note="MER81 repeat: matches 29. .68 of consensus"
23941. .24238
repeat_region /note="AluSx repeat: matches 1. .298 of consensus"
24239. .24299
repeat_region /note="MER81 repeat: matches 68. .114 of consensus"
24300. .24342
repeat_region /note="L2 repeat: matches 2704. .2746 of consensus"
24659. .24792
repeat_region /note="MIR repeat: matches 6. .134 of consensus"
25494. .25748
repeat_region /note="L2 repeat: matches 2452. .2709 of consensus"
26052. .26349
repeat_region /note="AluB repeat: matches 12. .305 of consensus"
26843. .26880
repeat_region /note="L1PA15-16 repeat: matches -357. .-319 of consensus"
28067. .28136
repeat_region /note="MIR repeat: matches 61. .131 of consensus"
28210. .28305
repeat_region /note="MIR repeat: matches 50. .146 of consensus"
28493. .28607
repeat_region /note="MIR repeat: matches 28. .145 of consensus"
29321. .29572

Query Match 0.3%; Score 74; DB 92; Length 125937;
Best Local Similarity 100.0%; Pred. No. 7.7e-29;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12637 tttgtatttttagtagagacggggtttctccacgtgtggtcaggtggtctcaaacctct 12696
|||||
Db 39270 TTTTGTATTTTAGTAGACGGGGTTTCTCCACGTTGTCAGGTGCTCAAACTCCT 39211
QY 12697 gacctcaggtgatc 12710
|||||
Db 39210 GACCTCAGGTGATC 39197

RESULT 48
AL512423
LOCUS AL512423 135252 bp DNA HTG 24-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP4-539L13, *** SEQUENCING IN

ACCESSION AL512423
VERSION AL512423.2 GI:12043508
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens
ORGANISM human.
REFERENCE
AUTHORS Pavitt,R.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT
On Jan 5, 2001 this sequence version replaced gi:12001719.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj539L13
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 130150 bases at least Q40
Consensus quality: 131436 bases at least Q30
Consensus quality: 132672 bases at least Q20
Insert size: 134152; sum-of-contigs
Insert size: 145141; 4.6% error; agarose-fp
Quality coverage: 6.06x in Q20 bases; sum-of-contigs Quality
coverage: 5.85x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 17724: contig of 17724 bp in length
* 17725 17824: gap of 100 bp
* 17825 23731: contig of 5907 bp in length
* 23732 23831: gap of 100 bp
* 23832 26290: contig of 2459 bp in length
* 26291 26390: gap of 100 bp
* 26391 58904: contig of 32514 bp in length
* 58905 59004: gap of 100 bp
* 59005 61434: contig of 2430 bp in length
* 61435 61534: gap of 100 bp
* 61535 64114: contig of 2580 bp in length
* 64115 64214: gap of 100 bp
* 64215 76613: contig of 12399 bp in length
* 76614 76713: gap of 100 bp
* 76714 79317: contig of 2604 bp in length
* 79318 79417: gap of 100 bp
* 79418 103390: contig of 23973 bp in length
* 103391 103490: gap of 100 bp
* 103491 117564: contig of 14074 bp in length
* 117565 117664: gap of 100 bp
* 117665 119820: contig of 2156 bp in length
* 119821 119920: gap of 100 bp
* 119921 135252: contig of 15332 bp in length.

FEATURES
source

Location/Qualifiers
1..135252
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP4-539L13"
/clone_lib="RPCI-4"
1..17724

misc_feature

```
/note="assembly_fragment:01329
fragment_chain:1
clone_end:SP6
vector_side:left"
17825..23731
/note="assembly_fragment:01207
fragment_chain:1"
23832..26290
/note="assembly_fragment:02174
fragment_chain:2"
26391..58904
/note="assembly_fragment:02093
fragment_chain:2"
59005..61434
/note="assembly_fragment:00176"
61535..64114
/note="assembly_fragment:00188"
64215..76613
/note="assembly_fragment:00640"
76714..79317
/note="assembly_fragment:00819"
79418..103390
/note="assembly_fragment:01300"
103491..117564
/note="assembly_fragment:01409"
117665..119820
/note="assembly_fragment:01567"
119921..135252
/note="assembly_fragment:02053"
33087 a 33936 c 34166 g 32963 t 1100 others

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Query Match          0.3%; Score 74; DB 81; Length 135252;
Best Local Similarity 100.0%; Pred. No. 7.7e-29;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12637 tttgtatttagtagacggggtttctccacgtgttcaggtggtctcaaaactct 12696
|||||
Db 75836 TTTTGTATTTTATGATAGACGGGGTTTCTCCACGTTGTCAGGCTGCTCAAACTCCT 75895
|||||

QY 12697 gacctcagtgatc 12710
|||||
Db 75896 GACCTCAGGTGATC 75909

RESULT 49
AC026323/c
LOCUS
DEFINITION
Homo sapiens chromosome 3 clone RP11-129P2, WORKING DRAFT SEQUENCE,
16 unordered pieces.
AC026323
VERSION
AC026323.9 GI:11094491
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 155190)
Muzny,D.M., Adams,C., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Blum,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
```

```

* 99705 99804: gap of unknown length
* 99805 109357: contig of 9553 bp in length
* 109358 109457: gap of unknown length
* 109458 118740: contig of 9283 bp in length
* 118741 118840: gap of unknown length
* 118841 126400: contig of 7560 bp in length
* 126401 126500: gap of unknown length
* 126501 134934: contig of 8494 bp in length
* 134935 135094: gap of unknown length
* 135095 139557: contig of 4463 bp in length
* 139558 139657: gap of unknown length
* 139658 143382: contig of 3725 bp in length
* 143383 143482: gap of unknown length
* 143483 147446: contig of 3964 bp in length
* 147447 147546: gap of unknown length
* 147547 150312: contig of 2766 bp in length
* 150313 150412: gap of unknown length
* 150413 152837: contig of 2425 bp in length
* 152838 152937: gap of unknown length
* 152938 154035: contig of 1098 bp in length
* 154036 154135: gap of unknown length
* 154136 155190: contig of 1055 bp in length.
FEATURES             Location/Qualifiers
     1..155190
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /chromosome="3"
     /clone="RP11-129P2"
BASE COUNT  47836 a 29970 c 29096 g 46758 t 1530 others
ORIGIN

Query Match      0.3%; Score 74; DB 70; Length 155190;
Best Local Similarity 100.0%; Pred. No. 7.9e-29;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10475 gcaccaccatgccagctaattttgtatttttaagtagacaggggtttccaccatgtt 10534
          |||||||
Db 144154 GCACACCATGCCAGCTAATTTGTATTATTTTGTAGACAGCGGGGTTTCCACCATGTT 144095

QY 10535 ggcacgagtgctct 10548
          |||||||
Db 144094 GGCACGAGTGGTCT 144081

RESULT 50
AC023995/c
LOCUS          AC023995 159057 bp DNA HTG 12-MAR-2000
DEFINITION    Homo sapiens chromosome 6 clone RP11-111D3 map 6, WORKING DRAFT
SEQUENCE     26 unordered pieces.
ACCESSION     AC023995
VERSION       AC023995.2 GI:7229926
KEYWORDS      HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 159057)
AUTHORS       Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
               Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
               Bouckgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
               Choephel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
               DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
               Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
               Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
               Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
               Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
               Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
               Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,

```

```

McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisani,C., Poillara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:7008900.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7040
Center clone name: l1l_D_3
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 142434 bases at least Q40
Consensus quality: 150924 bases at least Q30
Consensus quality: 154385 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 156557; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 1355: contig of 1355 bp in length
* 1356 1455: gap of 100 bp
* 1456 2614: contig of 1159 bp in length
* 2615 2714: gap of 100 bp
* 2715 4130: contig of 1416 bp in length
* 4131 4230: gap of 100 bp
* 4231 5925: contig of 1695 bp in length
* 5926 6025: gap of 100 bp
* 6026 7675: contig of 1650 bp in length
* 7676 7775: gap of 100 bp
* 7776 9827: contig of 2052 bp in length
* 9828 9927: gap of 100 bp
* 9928 11670: contig of 1743 bp in length
* 11671 11770: gap of 100 bp
* 11771 12351: contig of 581 bp in length
* 12352 12451: gap of 100 bp
* 12452 14390: contig of 1939 bp in length
* 14391 14490: gap of 100 bp
* 14491 17367: contig of 2877 bp in length
* 17368 17467: gap of 100 bp
* 17468 21354: contig of 3887 bp in length
* 21355 21454: gap of 100 bp
* 21455 24095: contig of 2641 bp in length
* 24096 24195: gap of 100 bp
* 24196 29114: contig of 4919 bp in length
* 29115 29214: gap of 100 bp
* 29215 35715: contig of 6501 bp in length
* 35716 35815: gap of 100 bp

```

TITLE

JOURNAL

COMMENT

* 35816 44434: contig of 8619 bp in length
* 44435 44534: gap of 100 bp
* 44535 52098: contig of 7564 bp in length
* 52099 52198: gap of 100 bp
* 52199 60217: contig of 8019 bp in length
* 60218 60317: gap of 100 bp
* 60318 68951: contig of 8634 bp in length
* 68952 69051: gap of 100 bp
* 69052 75496: contig of 6445 bp in length
* 75497 75596: gap of 100 bp
* 75597 83324: contig of 7728 bp in length
* 83325 83424: gap of 100 bp
* 83425 93400: contig of 9376 bp in length
* 93401 93500: gap of 100 bp
* 93501 102521: contig of 9021 bp in length
* 102522 102621: gap of 100 bp
* 102622 112785: contig of 10164 bp in length
* 112786 112885: gap of 100 bp
* 112886 124532: contig of 11647 bp in length
* 124533 124632: gap of 100 bp
* 124633 140780: contig of 16148 bp in length
* 140781 140880: gap of 100 bp
* 140881 159057: contig of 18177 bp in length.

FEATURES

source

Location/Qualifiers
1. 159057
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-111D3"
/clone_lib="RPC1-11 Human Male BAC"
1. 1355
/note="assembly_fragment"
1456..2614
/note="assembly_fragment"
2715..4130
/note="assembly_fragment"
4231..5925
/note="assembly_fragment"
6026..7675
/note="assembly_fragment"
7776..9827
/note="assembly_fragment"
9928..11670
/note="assembly_fragment"
11771..12351
/note="assembly_fragment"
clone_end:SP6
vector_side:right
12452..14390
/note="assembly_fragment"
14491..17367
/note="assembly_fragment"
17468..21354
/note="assembly_fragment"
21455..24095
/note="assembly_fragment"
clone_end:I7
vector_side:right
24196..29114
/note="assembly_fragment"
29215..35715
/note="assembly_fragment"
35816..44434
/note="assembly_fragment"
44535..52098
/note="assembly_fragment"
52199..60217
/note="assembly_fragment"
60318..68951
/note="assembly_fragment"
69052..75496
/note="assembly_fragment"

misc_feature 75597..83324
/note="assembly_fragment"
misc_feature 83425..93400
/note="assembly_fragment"
misc_feature 93501..102521
/note="assembly_fragment"
misc_feature 102622..112785
/note="assembly_fragment"
misc_feature 112886..124532
/note="assembly_fragment"
misc_feature 124633..140780
/note="assembly_fragment"
misc_feature 140881..159057
/note="assembly_fragment"
BASE COUNT 50480 a 28199 c 27784 g 50094 t 2500 others
ORIGIN

Query Match 0.3%; Score 74; DB 68; Length 159057;
Best Local Similarity 100.0%; Pred. No. 7.9e-29;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4707 cagctaattttttgtatttttagtagagacgggggtttaccatgtgtgccagagatagct 4766
|||||
Db 97410 CAGCTAATTTTGTATTTTAGTAGACGGGGTTTCACCATGTGGCCAGGATAGTCT 97351
|||||

Qy 4767 gcatctcttgacct 4780
|||||

Db 97350 CGATCTCTTGACCT 97337
|||||

RESULT 51

AC021346/C

LOCUS AC021346 170795 bp DNA HTG 09-SEP-2000
DEFINITION Homo sapiens clone RP11-24K15, WORKING DRAFT SEQUENCE, 12 unordered
pieces.

ACCESSION AC021346

VERSION AC021346.3 GI:10045520

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 170795)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

AUTHORS Homo sapiens, clone RP11-24K15

TITLE Unpublished

JOURNAL 2 (bases 1 to 170795)

REFERENCE 1 (bases 1 to 170795)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

Direct Submission

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Sep 9, 2000 this sequence version replaced gi:6806856.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

TITLE

JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4485
Center clone name: 24_K_15
----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 163380 bases at least Q40
Consensus quality: 167262 bases at least Q30
Consensus quality: 168687 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 169695; sum-of-contigs
Quality coverage: 5.3 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 37473: contig of 37473 bp in length
37474 37573: gap of 100 bp
37574 3817: contig of 1144 bp in length
3817 38817: gap of 100 bp
38818 39840: contig of 1023 bp in length
39841 39940: gap of 100 bp
39941 41368: contig of 1428 bp in length
41369 41468: gap of 100 bp
41469 43847: contig of 2379 bp in length
43848 43947: gap of 100 bp
43948 47918: contig of 3971 bp in length
47919 48018: gap of 100 bp
48019 59010: contig of 10992 bp in length
59011 59110: gap of 100 bp
59111 101273: contig of 42163 bp in length
101274 101373: gap of 100 bp
101374 113913: contig of 12540 bp in length
113914 114013: gap of 100 bp
114014 126871: contig of 12858 bp in length
126872 126971: gap of 100 bp
126972 160480: contig of 33509 bp in length
160481 160580: gap of 100 bp
160581 170795: contig of 10215 bp in length.

FEATURES
Source

1. 170795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-24K15"
/clone_lib="RPC1-11 Human Male BAC"

misc_feature

1. 37473
/note="assembly_fragment"
clone_end:SP6
vector_side:left"

misc_feature

37574. 38717

misc_feature

38818. 39840

misc_feature

39941. 41368

misc_feature

41469. 43847

misc_feature

43948. 47918

misc_feature

48019. 59010

misc_feature

59111. 101273

misc_feature
101374. 113913
/note="assembly_fragment"
misc_feature
114014. 126871
/note="assembly_fragment"
misc_feature
126972. 160480
/note="assembly_fragment"
misc_feature
160581. 170795
/note="assembly_fragment"
vector_side:right"

BASE COUNT 54926 a 33714 c 33988 g 47062 t 1105 others
ORIGIN

Query Match 0.3%; Score 74; DB 66; Length 170795;

Best Local Similarity 100.0%; Pred. No. 8e-29; Indels 0; Gaps 0;
Matches 74; Conservative 0; Mismatches 0;

QY 12637 ttgttatttttagtagacggggtttctcacgttggtcagcggtgtctcaaacctcct 12696
|||||

Db 156947 TTTTGTATTTTAGTAGACGGGGTTTCTCCACGTTGTCAGGCTGTCCTCAAACTCCT 156888
|||||

QY 12697 gacctcaggtgatc 12710
|||||

Db 156887 GACCTCAGGTGATC 156874
|||||

RESULT 52

AC022735/c

LOCUS

AC022735 176731 bp DNA HTG 03-FEB-2001
Homo sapiens chromosome 15 clone RP11-519P13 map 15, WORKING DRAFT
SEQUENCE, 5 unordered pieces.

ACCESSION

AC022735

VERSION

AC022735.4 GI:12658120

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 176731)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

JOURNAL

REFERENCE

2 (bases 1 to 176731)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,

Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,

Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,

Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,

Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,

Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,

McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,

Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,

Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,

Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,

Zimmer,A. and Zody,M.

Direct Submission

TITLE

JOURNAL

COMMENT

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 3, 2001 this sequence version replaced gi:7249212.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L597
 Center clone name: 519_P_13

----- Summary Statistics
 Sequencing vector: M13; M7815; 57% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 174754 bases at least Q40
 Consensus quality: 175799 bases at least Q30
 Consensus quality: 176185 bases at least Q20
 Insert size: 158000; agarose-fp
 Insert size: 176331; sum-of-contigs
 Quality coverage: 8.8 in Q20 bases; agarose-fp
 Quality coverage: 7.9 in Q20 b.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

* 1 2067: contig of 2067 bp in length
 * 2168 86423: gap of 100 bp
 * 86424 86523: contig of 84256 bp in length
 * 86524 105541: gap of 100 bp
 * 105542 105641: contig of 19018 bp in length
 * 105642 160623: gap of 100 bp
 * 160624 160723: contig of 54982 bp in length
 * 160724 176731: gap of 100 bp
 * 160724 176731: contig of 16008 bp in length.

FEATURES

source

1. .176731
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15"

/clone="RP11-519P13"
 /clone_lib="RPC1-11 Human Male BAC"

1. .2067

/note="assembly_fragment"
 clone_end:SP6
 vector_side:left"

2168. 86423

/note="assembly_fragment"

86524. 105541

/note="assembly_fragment"

105642. 160623

/note="assembly_fragment"

160724. 176731

/note="assembly_fragment"
 clone_end:T7
 vector_side:right"

BASE COUNT 55976 a 35801 c 34289 g 50265 t 400 others
 ORIGIN

Query Match

Best Local Similarity 100.0%; Score 74; DB 67; Length 176731;

Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17360 tttttgagcggaggttactcttctgtccaggtcggagtgcaatggcgtgatctcagc 17419

|||||

Db 29699 TTTTGTGAGACGGAGTTTCACTCTGTGTGCCAGGCTGGAGTGCATATCTCTCAGC 29640

|||||

QY 17420 tcactgcaacctcc 17433

|||||

Db 29639 TCACTGCACCTCC 29626

RESULT 53

AC025002

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC025002 200484 bp DNA HTG 14-JUN-2000
 Homo sapiens chromosome 1 clone RP11-204L3 map 1, *** SEQUENCING IN
 PROGRESS ***, 81 unordered pieces.

AC025002

AC025002.3

GI:8516086

HTG: HTGS_PHASE1.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 200484)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 1, clone RP11-204L3

Unpublished

2 (bases 1 to 200484)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,

Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,

Galagan,J., Gardyna,S., Glende,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,

Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,

Meidirm,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,

Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigillo,J.,

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,

Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7675

Center clone name: 204_L_3

* NOTE: This is a 'working draft' sequence. It currently

* consists of 81 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1048: contig of 1048 bp in length

* 1049 1148: gap of 100 bp

* 1149 2470: contig of 1322 bp in length

* 2471 2570: gap of 100 bp

* 2571 3643: contig of 1073 bp in length

* 3644 3743: gap of 100 bp

* 3744 4980: contig of 1237 bp in length

* 4981 5080: gap of 100 bp

* 5081 6352: contig of 1272 bp in length

* 6353 6452: gap of 100 bp

* 6453 7549: contig of 1097 bp in length

* 7550 7649: gap of 100 bp

* 7650 7649: gap of 100 bp

* 7650 7649: gap of 100 bp

* 7650 7649: gap of 100 bp

* 7650 7649: gap of 100 bp

* 7650 7649: gap of 100 bp

* 7650 7649: gap of 100 bp

* 7650 7649: gap of 100 bp

```
* 7650 8977: contig of 1328 bp in length
* 8978 9077: gap of 100 bp
* 9078 10103: contig of 1026 bp in length
* 10104 10203: gap of 100 bp
* 10204 11362: contig of 1159 bp in length
* 11363 11462: gap of 100 bp
* 11463 12821: contig of 1359 bp in length
* 12822 12921: gap of 100 bp
* 12922 14007: contig of 1086 bp in length
* 14008 14107: gap of 100 bp
* 14108 15283: contig of 1176 bp in length
* 15284 15383: gap of 100 bp
* 15384 16399: contig of 1016 bp in length
* 16400 16499: gap of 100 bp
* 16500 17804: contig of 1305 bp in length
* 17805 17904: gap of 100 bp
* 17905 19176: contig of 1272 bp in length
* 19177 19276: gap of 100 bp
* 19277 20384: contig of 1108 bp in length
* 20385 20484: gap of 100 bp
* 20485 21689: contig of 1205 bp in length
* 21690 21789: gap of 100 bp
* 21790 23191: contig of 1402 bp in length
* 23192 23291: gap of 100 bp
* 23292 24513: contig of 1222 bp in length
* 24514 24613: gap of 100 bp
* 24614 26135: contig of 1522 bp in length
* 26136 26235: gap of 100 bp
* 26236 27373: contig of 1138 bp in length
* 27374 27473: gap of 100 bp
* 27474 28723: contig of 1250 bp in length
* 28724 28823: gap of 100 bp
* 28824 30217: contig of 1394 bp in length
* 30218 30317: gap of 100 bp
* 30318 32056: contig of 1739 bp in length
* 32057 32156: gap of 100 bp
* 32157 33818: contig of 1662 bp in length
* 33819 33918: gap of 100 bp
* 33919 35156: contig of 1238 bp in length
* 35157 35256: gap of 100 bp
* 35257 36712: contig of 1456 bp in length
* 36713 36812: gap of 100 bp
* 36813 38379: contig of 1567 bp in length
* 38380 38479: gap of 100 bp
* 38480 40190: contig of 1711 bp in length
* 40191 40290: gap of 100 bp
* 40291 41780: contig of 1490 bp in length
* 41781 41880: gap of 100 bp
* 41881 44053: contig of 2173 bp in length
* 44054 44153: gap of 100 bp
* 44154 45474: contig of 1321 bp in length
* 45475 45574: gap of 100 bp
* 45575 46821: contig of 1247 bp in length
* 46822 46921: gap of 100 bp
* 46922 47302: contig of 381 bp in length
* 47303 47402: gap of 100 bp
* 47403 49323: contig of 1921 bp in length
* 49324 49423: gap of 100 bp
* 49424 51140: contig of 1717 bp in length
* 51141 51240: gap of 100 bp
* 51241 53099: contig of 1859 bp in length
* 53100 53199: gap of 100 bp
* 53200 54631: contig of 1432 bp in length
* 54632 54731: gap of 100 bp
* 54732 55751: contig of 1020 bp in length
* 55752 55851: gap of 100 bp
* 55852 57470: contig of 1619 bp in length
* 57471 57570: gap of 100 bp
* 57571 60182: contig of 2612 bp in length
* 60183 60282: gap of 100 bp
* 60283 63065: contig of 2783 bp in length
* 63066 63165: gap of 100 bp
* 63166 64443: contig of 1278 bp in length
```

```
* 64444 64543: gap of 100 bp
* 64544 66043: contig of 1500 bp in length
* 66044 66143: gap of 100 bp
* 66144 68623: contig of 2480 bp in length
* 68624 68723: gap of 100 bp
* 68724 70122: contig of 1399 bp in length
* 70123 70222: gap of 100 bp
* 70223 72407: contig of 2185 bp in length
* 72408 72507: gap of 100 bp
* 72508 74635: contig of 2128 bp in length
* 74636 74735: gap of 100 bp
* 74736 77554: contig of 2819 bp in length
* 77555 77654: gap of 100 bp
* 77655 80277: contig of 2623 bp in length
* 80278 80377: gap of 100 bp
* 80378 81798: contig of 1421 bp in length
* 81799 81898: gap of 100 bp
* 81899 84126: contig of 2228 bp in length
* 84127 84226: gap of 100 bp
* 84227 86794: contig of 2568 bp in length
* 86795 86894: gap of 100 bp
* 86895 90311: contig of 3417 bp in length
* 90312 90411: gap of 100 bp
* 90412 92255: contig of 1844 bp in length
* 92256 92355: gap of 100 bp
* 92356 94570: contig of 2215 bp in length
* 94571 94670: gap of 100 bp
* 94671 97163: contig of 2493 bp in length
* 97164 97263: gap of 100 bp
* 97264 99724: contig of 2461 bp in length
* 99725 99824: gap of 100 bp
* 99825 103151: contig of 3327 bp in length
* 103152 103251: gap of 100 bp
* 103252 106370: contig of 3119 bp in length
* 106371 106470: gap of 100 bp
* 106471 108888: contig of 2418 bp in length
* 108889 108988: gap of 100 bp
* 108989 111514: contig of 2526 bp in length
* 111515 111614: gap of 100 bp
* 111615 114360: contig of 2746 bp in length
* 114361 114460: gap of 100 bp
* 114461 118269: contig of 3809 bp in length
* 118270 118369: gap of 100 bp
* 118370 122359: contig of 3990 bp in length
* 122360 122459: gap of 100 bp
* 122460 125669: contig of 3210 bp in length
* 125670 125769: gap of 100 bp
* 125770 129306: contig of 3537 bp in length
* 129307 129406: gap of 100 bp
* 129407 132984: contig of 3578 bp in length
* 132985 133084: gap of 100 bp
* 133085 137360: contig of 4276 bp in length
* 137361 137460: gap of 100 bp
* 137461 140476: contig of 3016 bp in length
* 140477 140576: gap of 100 bp
```

Query Match 0.3%; Score 74; DB 69; Length 200484;

Best Local Similarity 100.0%; Pred. No. 8.2e-29; Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12637 tttgtatttttagtagagacgggtttctccacgttggtcaggtgtctcaaacctcct 12696

|||||

Db 164464 TTTTGTATTTTATTAGTAGACGAGGGGTTTCTCCAGGTGGTCTCAAACTCCT 164523

QY 12697 gacctcaggtgac 12710

|||||

Db 164524 GACCTCAGGTGATC 164537

RESULT 54

AC010553

LOCUS

DEFINITION

AC010553 85543 bp DNA

Homo sapiens chromosome 16 clone RP11-59D8, complete sequence.

PRI

28-JUN-2000

AC010553
 AC010553.6 GI:8778952
 HTG
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 85543)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 85543)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 85543)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (28-JUN-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Jun 28, 2000 this sequence version replaced gi:8576035.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center
 www.shgc.stanford.edu
 Quality: Phrap Quality >=40 99.4% of Sequence;
 Estimated Total Number of Errors is 0.5.
 STS Content:
 SHGC-33776 G29491
 WI-16772 G23176.

FEATURES
 Location/Qualifiers
 1..85543
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-59D8"
 BASE COUNT 23468 a 19951 c 19917 g 22207 t
 ORIGIN
 1..85543
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-59D8"
 BASE COUNT 23468 a 19951 c 19917 g 22207 t
 ORIGIN

Query Match 0.3%; Score 73; DB 87; Length 85543;
 Best Local Similarity 100.0%; Pred. No. 2.6e-28;
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17366 agacgaggttcactcttctccaggctgagtgcaatgcgctgactcagctcaatg 17425
 |||||
 DB 27620 AGACGAGGTTTCACTCTCTGTCGCCAGGCTGGAGTGCAATGGCTGATCTCAGCTCACATG 27679
 |||||
 QY 17426 caactccgcctc 17438
 |||||
 DB 27680 CAACCTCCGCTC 27692

RESULT 55
 AL360232
 LOCUS Human DNA sequence from clone RP11-374118 on chromosome 6, complete
 DEFINITION sequence.
 ACCESSION AL360232
 VERSION AL360232.24 GI:13625025
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 157784)
 AUTHORS Lawlor, S.
 TITLE Direct Submission
 JOURNAL Submitted (13-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 COMMENT On Apr 14, 2001 this sequence version replaced gi:13396590.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >=30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at
<http://www.sanger.ac.uk/projects/c-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 RP11-374118 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACE3.6
 IMPORTANT: This sequence is not the entire insert of clone RP11-374118. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
 The true left end of clone RP11-374118 is at 1 in this sequence.
 The true left end of clone RP11-380M3 is at 157685 in this sequence.
 The true right end of clone RP11-30B21 is at 48032 in this sequence.

FEATURES
 Location/Qualifiers
 1..157784
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="RP11-374118"
 /clone_lib="RPCI-11.2"
 2165..2605
 /note="MLTID repeat: matches 19..501 of consensus"
 repeat_region 4231..4286
 /note="14 copies 4 mer tctc 76% conserved"
 repeat_region 4595..4630
 /note="18 copies 2 mer aa 80% conserved"
 repeat_region 4705..4927
 /note="L2 repeat: matches 2511..2720 of consensus"
 repeat_region 5245..5284
 /note="L2 repeat: matches 2705..2750 of consensus"
 repeat_region 6066..8377
 /note="L1MA5A repeat: matches 4019..6283 of consensus"
 repeat_region 8401..8524
 /note="62 copies 2 mer at 68% conserved"
 repeat_region 8411..8522
 /note="28 copies 4 mer atat 70% conserved"
 repeat_region 9401..11199
 /note="L1PA5 repeat: matches 4337..6141 of consensus"
 repeat_region 11887..12021
 /note="MER20 repeat: matches 81..218 of consensus"
 repeat_region 12705..12792
 /note="MIR repeat: matches 106..196 of consensus"
 repeat_region 13252..13442
 /note="L1ME3 repeat: matches 5731..5929 of consensus"
 repeat_region 15051..15271
 /note="MIR repeat: matches 7..251 of consensus"
 repeat_region 15273..15540
 /note="MER63B repeat: matches 1..272 of consensus"
 repeat_region 15926..16160
 /note="AluJo repeat: matches 49..285 of consensus"
 repeat_region 16688..16857
 /note="MIR repeat: matches 92..261 of consensus"
 repeat_region 17433..17598

repeat_region /note="LIPAA repeat: matches 5981. .6146 of consensus"
18140. .18446
repeat_region /note="AluY repeat: matches 1. .307 of consensus"
18733. .19120
repeat_region /note="LIME repeat: matches 875. .1259 of consensus"
19140. .19377
repeat_region /note="LIMA7 repeat: matches 6041. .6284 of consensus"
19381. .19416
repeat_region /note="RNA-Gly-GGC repeat: matches 1. .36 of consensus"
19417. .19822
repeat_region /note="LIMA7 repeat: matches 5564. .6059 of consensus"
19821. .20173
repeat_region /note="LIME repeat: matches 424. .865 of consensus"
21920. .22030
repeat_region /note="MIR repeat: matches 36. .147 of consensus"
22044. .23086
repeat_region /note="LIPA7 repeat: matches 4518. .5567 of consensus"
23087. .23637
repeat_region /note="LIPA7 repeat: matches 5564. .6145 of consensus"
25410. .25529
repeat_region /note="MIR repeat: matches 40. .157 of consensus"
25873. .26155
repeat_region /note="AluJo repeat: matches 1. .288 of consensus"
26575. .26610
repeat_region /note="18 copies 2 mer aa 80% conserved"
26638. .26927
repeat_region /note="AluSp repeat: matches 5. .299 of consensus"
27668. .27868
repeat_region /note="LIPAS repeat: matches 5963. .6163 of consensus"
28182. .28472
repeat_region /note="THELC repeat: matches 84. .371 of consensus"
28473. .28573
repeat_region /note="LIMAA repeat: matches 5802. .5912 of consensus"
28581. .28624
repeat_region /note="22 copies 2 mer ac 90% conserved"
28635. .28947
repeat_region /note="LIMAA repeat: matches 5468. .5782 of consensus"
31677. .31741
repeat_region /note="MER63B repeat: matches 372. .436 of consensus"
31742. .32147
repeat_region /note="LIMAA repeat: matches 5655. .6057 of consensus"
32324. .32544
repeat_region /note="MER63B repeat: matches 3. .237 of consensus"
32833. .33066
repeat_region /note="AluX repeat: matches 49. .282 of consensus"
34193. .34220
repeat_region /note="7 copies 4 mer acac 100% conserved"
34227. .34262
repeat_region /note="18 copies 2 mer gt 91% conserved"
34230. .34261
repeat_region /note="8 copies 4 mer tgtg 93% conserved"
35911. .35966
repeat_region /note="14 copies 4 mer acac 91% conserved"
35917. .35966
repeat_region /note="25 copies 2 mer ac 94% conserved"
37344. .37574
repeat_region /note="L2 repeat: matches 2453. .2746 of consensus"
38216. .38508
repeat_region /note="LIMD2 repeat: matches 6055. .6341 of consensus"
39051. .39360
repeat_region /note="AluJb repeat: matches 1. .297 of consensus"
39848. .40049
repeat_region /note="MIR repeat: matches 29. .240 of consensus"
40309. .40348
repeat_region /note="20 copies 2 mer tg 95% conserved"
41101. .41242
repeat_region /note="MER69A repeat: matches 19. .168 of consensus"
41243. .41302
repeat_region /note="LIPA2 repeat: matches 6087. .6146 of consensus"
42093. .42398
repeat_region /note="AluJo repeat: matches 1. .306 of consensus"
44475. .44692
repeat_region /note="MIR repeat: matches 18. .252 of consensus"

45432. .45671
repeat_region /note="MIR repeat: matches 10. .255 of consensus"
45724. .46025
repeat_region /note="AluJb repeat: matches 1. .303 of consensus"
47127. .47397
repeat_region /note="AluJo repeat: matches 1. .277 of consensus"
47398. .47460
repeat_region /note="21 copies 3 mer aat 84% conserved"
47581. .47671
repeat_region /note="MER91A repeat: matches 35. .127 of consensus"
48328. .48607
repeat_region /note="LIMA3 repeat: matches 6016. .6300 of consensus"
49737. .50021
repeat_region /note="AluJb repeat: matches 1. .309 of consensus"
51444. .51497
repeat_region /note="27 copies 2 mer aa 72% conserved"
51938. .52144
repeat_region /note="MIR repeat: matches 38. .262 of consensus"
54096. .54112
repeat_region /note="MIR repeat: matches 247. .262 of consensus"
54113. .54154
repeat_region /note="L2 repeat: matches 2649. .2690 of consensus"
54155. .54181
repeat_region /note="MIR repeat: matches 173. .247 of consensus"
56527. .57321
repeat_region /note="LIPAB repeat: matches 5367. .6163 of consensus"
57460. .57763
repeat_region /note="AluSq repeat: matches 1. .305 of consensus"
57984. .58081
repeat_region /note="LIMC/D repeat: matches 5324. .5424 of consensus"
58501. .58927
repeat_region /note="LIME repeat: matches 5320. .5745 of consensus"
59346. .59645
repeat_region /note="LIM4 repeat: matches 4534. .4845 of consensus"
59646. .60002
repeat_region /note="MLT1A1 repeat: matches 1. .365 of consensus"
60003. .60021
repeat_region

Query Match 0.3%; Score 73; DB 90; Length 157784;
Best Local Similarity 100.0%; Pred. No. 2.9e-28;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11112 ccaccacacttggttaattttgttttagtagagacgggtttccacatgttgcca 11171
|||||
Db 136792 CCACCACACTGGCTAATTTTGTGTTTGTAGTAGACGGGGTTTCCACCATGTGGCCA 136851
|||||

QY 11172 gctggtcttgaa 11184
|||||
Db 136852 GGCTGCTTGAA 136864
|||||

RESULT 56
AC013245 163884 bp DNA HTG 03-OCT-2000
LOCUS Homo sapiens chromosome 12 clone RP11-329010, WORKING DRAFT
DEFINITION SEQUENCE, 7 unordered pieces.
ACCESSION AC013245
VERSION AC013245.18 GI:10445256
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163884)
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-oshman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbaria, J.,
Benton, J., Brieva, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 87896: contig of 87896 bp in length
* 87897 87996: gap of 100 bp
* 87997 137489: contig of 49493 bp in length
* 137490 137589: gap of 100 bp
* 137590 168074: contig of 30485 bp in length
* 168075 168174: gap of 100 bp
* 168175 170782: contig of 2608 bp in length
* 170783 170882: gap of 100 bp
* 170883 193830: contig of 22948 bp in length.

FEATURES

Source
1..193830
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-309F17"
/clone_lib="RPC1-11.2"
1..87896
/note="assembly:fragment:01107
fragment_chain:1"
87997..137489
/note="assembly:fragment:01732
fragment_chain:1"
137590..168074
/note="assembly:fragment:01468
fragment_chain:1"
168175..170782
/note="assembly:fragment:02730
fragment_chain:1"
170883..193830
/note="assembly:fragment:00861
fragment_chain:1
clone_end:T7
vector_side:right"
57880 a 38659 c 38280 g 58611 t 400 others

BASE COUNT 57880 a 38659 c 38280 g 58611 t 400 others

ORIGIN

Query Match 0.3%; Score 72; DB 82; Length 193830;
Best Local Similarity 100.0%; Pred. No. 1.1e-27;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4709 gctaatTTTTgtatttttagtagagacgggtttccaccatgttgccagatagctcg 4768
|||||
Db 130505 GCTAATTTTTGTATTTTGTAGTAGACGGGTTTCACCATGTGGCCAGGATAGTCTCG 130564
|||||

QY 4769 atctctgacct 4780
|||||
Db 130565 ATCTCTTGACCT 130576
|||||

RESULT 61
AC010900/c
LOCUS
DEFINITION Homo sapiens chromosome UNK clone RP11-544H14, WORKING DRAFT
SEQUENCE, 3 unordered pieces.
AC010900
AC010900.11 GI:13431240
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 208248)
Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 208248)
Waterston, R.H.
AUTHORS Direct Submission
TITLE Submitted (25-SEP-1999) Genome Sequencing Center, Washington

COMMENT

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 22, 2001 this sequence version replaced gi:12740570.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0544H14
----- Summary Statistics -----
Sequencing vector: M13; 45%
Sequencing vector: plasmid; 36%
Chemistry: Dye-terminator ET; 38% of reads
Chemistry: Dye-terminator Big Dye; 43% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 207248 bases at least Q40
Consensus quality: 207697 bases at least Q30
Consensus quality: 207892 bases at least Q20
Insert size: 206000; agarose-fp
Insert size: 207997; sum-of-contigs
Quality coverage: 7.85 in Q20 bases; agarose-fp
Quality coverage: 7.77 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 51: contig of 51 bp in length
* 52 151: gap of unknown length
* 152 12457: contig of 12306 bp in length
* 12458 12557: gap of unknown length
* 12558 208248: contig of 195691 bp in length.

FEATURES

Source
1..208248
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-544H14"
1..51
misc_feature
/note="assembly_name:Contig20"
152..12457
misc_feature
/note="assembly_name:Contig32"
12558..208248
misc_feature
/note="assembly_name:Contig33"
BASE COUNT 61720 a 41384 c 41137 g 63807 t 200 others

ORIGIN

Query Match 0.3%; Score 72; DB 61; Length 208248;
Best Local Similarity 100.0%; Pred. No. 1.1e-27;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18972 ggagatcgagaccatctgctgaacatgatgaacccgctctactataatacaaaaa 19031
|||||
Db 162487 GGAGATCGAGACCATCTCGTGAACATGATGAACCCCTCTCTACTAAAAATACAAAAA 162428
|||||

QY 19032 ttactctggcgt 19043
|||||
Db 162427 TTACTCTGGCGCT 162416
|||||

RESULT 62
AC004802/c
LOCUS
DEFINITION Homo sapiens 12p13.3 RPC14-773N5 (Roswell Park Cancer Institute
Human PAC library) complete sequence.
AC004802
AC004802.1 GI:3406033

AC004802 57304 bp DNA PRI 17-SEP-1998
Homo sapiens 12p13.3 RPC14-773N5 (Roswell Park Cancer Institute
Human PAC library) complete sequence.

AC004802
AC004802.1 GI:3406033


```

KEYWORDS
SOURCE
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Muzny,D., Arenson,A.D., Adams,C., Bunac,C., Carvelli,K., Chang,J.,
              Chacko,J., Chen,J., Ding,Y., Dugan,S., Durbin,J., Forcum,J.,
              Ganesh,R., Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M.,
              Hernandez,J., Jackson,L., Jin,S., Kampal,R., Karpathy,S., Kovar,C.,
              Lau,S., Leal,B., Lee,E., Li,Y., Lichtarge,O., Liu,W., Logan,O.,
              Lu,J., Ly,T., Marondel,I., Martinez,C., Merscher,S., Montgomery,K.,
              Oswal,G., Perez,L., Rashid,N.D., Renault,B., Rowland,K., Savage,L.,
              Scherer,S.E., Shen,H., Simon,M., Stovall,K., Timms,K.M., Todd,J.,
              Vo,Q., Williamson,A., Worley,K.C., Yu,W., Kucherlapati,R.,
              Nelson,D. and Gibbs,R.A.
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (06-JUN-1998) Molecular and Human Genetics, Baylor
              College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (08-AUG-1998) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
REFERENCE
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (16-SEP-1998) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
REFERENCE
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (17-SEP-1998) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
COMMENT
On Aug 8, 1998 this sequence version replaced gi:3337341.
Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by
Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui
Zhang.

Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
FEATURES
source
Location/Qualifiers
1..57304
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RPC14-773N5"
/chromosome="12p13.3"
misc_feature
1..347
/note="Overlaps with AC004804 bases 108534 - 108879"
/function="Overlap with adjacent clone AC004804"
misc_feature
complement(802..1201)
/note="Region: Similar to AAL49366"
repeat_region
complement(891..917)
/rpt_family="AT_rich"
repeat_region
2015..2141
/rpt_family="LIM4"
repeat_region
complement(2707..2837)
/rpt_family="FLAM_C"
/rpt_family="THE1B"
repeat_region
complement(3139..3168)
/rpt_family="AT_rich"
repeat_region
3238..3498
/rpt_family="L2"
repeat_region
3657..3789
/rpt_family="FLAM_C"
repeat_region
4090..4394
/rpt_family="AluSx"
repeat_region
complement(5467..5491)
/rpt_family="AT_rich"
repeat_region
5912..6014
/rpt_family="AluJo"
repeat_region
6015..6298
/rpt_family="AluSg"
repeat_region
6301..6471
/rpt_family="AluJo"
repeat_region
complement(6878..6905)
/rpt_family="AT_rich"
repeat_region
complement(7068..7360)
/rpt_family="AluSg"
repeat_region
complement(7361..7489)
/rpt_family="AluJo/FLAM"
repeat_region
7967..8012
/rpt_family="AT_rich"
repeat_region
8278..8562
/rpt_family="AluJo"
repeat_region
complement(9425..9706)
/rpt_family="AluSp"
repeat_region
10111..10397
/rpt_family="AluSx"
repeat_region
10905..10969
/rpt_family="(CA)n"
repeat_region
complement(11345..11399)
/rpt_family="L2"
repeat_region
11417..11717
/rpt_family="AluYa5"
repeat_region
complement(12136..12169)
/rpt_family="AT_rich"
repeat_region
complement(12251..12483)
/rpt_family="L2"
repeat_region
13351..14462
/rpt_family="TIGGER1"
repeat_region
14463..14644
/rpt_family="AluSg/x"
repeat_region
14666..15388
/rpt_family="TIGGER1"
repeat_region
15389..15687
/rpt_family="AluSg"
repeat_region
15688..16316
/rpt_family="TIGGER1"
repeat_region
18435..18472
/rpt_family="AT_rich"
repeat_region
complement(18738..19035)
/rpt_family="AluJo"
repeat_region
19844..19869
/rpt_family="POLY_A"
repeat_region
20155..20212
/rpt_family="MIR"
repeat_region
20871..20915
/rpt_family="AT_rich"
repeat_region
complement(21622..21981)
/rpt_family="LFR16A"
repeat_region
22442..22741
/rpt_family="AluSg"
repeat_region
22806..23098
/rpt_family="AluJb"
repeat_region
23214..23508
/rpt_family="AluSg"
repeat_region
24489..24742
/rpt_family="LIM4"
repeat_region
complement(24744..25102)
/rpt_family="THE1B"

```

repeat_region	complement(25103..25230)	Query Match	0.3%; Score 71; DB 85; Length 57304;
repeat_region	/rpt_family="THEIB-internal"	Best Local Similarity	100.0%; Pred. No. 3.3e-27;
repeat_region	25212..25431	Matches 71; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
repeat_region	/rpt_family="LIMB6"		
repeat_region	25592..25641		
repeat_region	/rpt_family="Alu"		
repeat_region	25642..25914		
repeat_region	/rpt_family="AluY"		
repeat_region	26584..26882		
repeat_region	/rpt_family="AluY"		
repeat_region	26883..26991		
repeat_region	/rpt_family="(TAAAA)n"		
repeat_region	26993..27173		
repeat_region	/rpt_family="MLTIF"		
repeat_region	27209..27509		
repeat_region	/rpt_family="AluSx"		
repeat_region	27564..27747		
repeat_region	/rpt_family="MLTIF"		
repeat_region	complement(28495..28556)		
repeat_region	/rpt_family="AT_rich"		
repeat_region	complement(29309..29356)		
repeat_region	/rpt_family="AT_rich"		
repeat_region	30328..30461		
repeat_region	/rpt_family="FLAM_C"		
repeat_region	complement(30500..30766)		
repeat_region	/rpt_family="AluSx"		
repeat_region	complement(31336..31395)		
repeat_region	/rpt_family="MIR"		
repeat_region	31701..32022		
repeat_region	/rpt_family="MER1B"		
repeat_region	32472..32794		
repeat_region	/rpt_family="AluSp"		
STs	32560..32701		
	/standard_name="G02122"		
	/note="STS1-CSRL-24g1-UA/CSRL-24g1-UZ"		
repeat_region	/db_xref="dbSTS:7608"		
repeat_region	complement(33007..33035)		
repeat_region	/rpt_family="AT_rich"		
repeat_region	33390..33517		
repeat_region	/rpt_family="L2"		
repeat_region	33693..33766		
repeat_region	/rpt_family="LIP"		
repeat_region	33872..34126		
repeat_region	/rpt_family="AluJb"		
repeat_region	34687..34986		
repeat_region	/rpt_family="AluSg"		
repeat_region	complement(35119..35151)		
repeat_region	/rpt_family="AT_rich"		
repeat_region	35327..35360		
QY 17379	ctcttgttccccagctggagtgcaatggcggtgatctcagctcactgcaacctcgcctc 17438		
Db 56463	CTCTTGTGCCCAGGTGGAGTGCAATGGCGGTGATCTCAGTCACTCAACCTCCGCTC 56404		
QY 17439	ccgggttcgaag 17449		
Db 56403	CCGGGTTCGAAG 56393		
RESULT 63			
AC083783/C			
LOCUS	AC083783	65916 bp	DNA
DEFINITION	Homo sapiens chromosome 17 clone RP11-462C21 map 17, LOW-PASS		HTG
	SEQUENCE SAMPLING.		
ACCESSION	AC083783		
VERSION	AC083783.2	GI:13194285	
KEYWORDS	HTG; HTGS_PHASE0.		
SOURCE	human.		
ORGANISM	Homo sapiens		

Chemistry: Dye-terminator Big Dye; 20% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 186330 bases at least Q40
Consensus quality: 186828 bases at least Q30
Consensus quality: 187104 bases at least Q20
Insert size: 195000; agarose-fp
Insert size: 187779; sum-of-contigs
Quality coverage: 7.15 in Q20 bases; agarose-fp
Quality coverage: 7.45 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N. But the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 3458: contig of 3458 bp in length
* 3459 3558: gap of unknown length
* 3559 8745: contig of 5187 bp in length
* 8746 8845: gap of unknown length
* 8846 20598: contig of 11753 bp in length
* 20599 20698: gap of unknown length
* 20699 30338: contig of 9640 bp in length
* 30339 30438: gap of unknown length
* 30439 44707: contig of 14269 bp in length
* 44708 44807: gap of unknown length
* 44808 57132: contig of 12325 bp in length
* 57133 57232: gap of unknown length
* 57233 72793: contig of 15561 bp in length
* 72794 72893: gap of unknown length
* 72894 96247: contig of 23354 bp in length
* 96248 96347: gap of unknown length
* 96348 122530: contig of 26183 bp in length
* 122531 122631: gap of unknown length
* 122631 188679: contig of 66049 bp in length.

FEATURES

source
1. .188679
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/clone="RP11-296K10"
1. .3458
/note="assembly_name:Contig24"
3559. .8745
/note="assembly_name:Contig25
clone_end:T7
vector_side:left"
8846. .20598
/note="assembly_name:Contig26"
20699. .30338
/note="assembly_name:Contig27"
30439. .44707
/note="assembly_name:Contig28"
44808. .57132
/note="assembly_name:Contig29
clone_end:SP6
vector_side:left"
57233. .72793
/note="assembly_name:Contig30"
72894. .96247
/note="assembly_name:Contig31"
96348. .122530
/note="assembly_name:Contig32"
122631. .188679
/note="assembly_name:Contig33"
51015 a 43483 c 42231 g 51048 t 902 others

BASE COUNT 51015 a 43483 c 42231 g 51048 t 902 others
ORIGIN

Query Match 0.3%; Score 71; DB 67; Length 188679;
Best Local Similarity, 100.0%; Pred. No. 3.9e-27;

Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17379 cttcttttccagctgaagtcgaatggcggtatctcaqctcactgcaacctccgctc 17438
|||||
Db 151717 CTCTTTGTTGCCAGGCTGGAGTGCAATGGCGTGATCTCAGTCACTGCAACCTCCGCTC 151658
|||||
QY 17439 ccgggttcaag 17449
|||||
Db 151657 CCGGGTTCAG 151647

RESULT 67
CNS01RHC

LOCUS
DEFINITION
Homo sapiens chromosome 14 clone R-736N17, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 192064)
Direct Submission
Submitted (07-DEC-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On Dec 9, 2000 this sequence version replaced gi:8247492.

Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr

COMMENT

IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc. . . even if efforts are made to eliminate these
contaminating sequences. The following BAC sequence is oriented
from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-45P15 (AC-ALL138976)
Downstream BAC (overlapping the SP6 end) : C-2538G10

Assembly program: Phrap; version 2.0
Quality coverage: 10.24x in Q20 bases; sum-of-contigs

Overall quality chart :

Range : bases
0 :
1 - 9 : 30
10 - 19 : 73
20 - 29 : 349
30 - 39 : 3955
40 - 49 : 6424
50 - 59 : 8119
60 - 69 : 19129
70 - 79 : 54514
80 - 89 : 99471
90 - 99 : 99471

Percentage of bases with a quality value >= 40 : 99 %.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
Location/Qualifiers
1. .192064
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/chromosome="14"
/clone="R-736N17"
/clone_lib="RPC1-11"
27321..27545
/note="matching EMBL:G04572
RHdb:RH53929
RHdb:RH3791
dbSTS:STS22999
Identified using the e-PCR software (G. Schuler)"
103114..103373
/note="matching EMBL:M92357
RHdb:RH17752
dbSTS:STS14628
Identified using the e-PCR software (G. Schuler)"
129887..130096
/note="matching EMBL:G32911
RHdb:RH67603
dbSTS:STS47530
Identified using the e-PCR software (G. Schuler)"
130031..130217
/note="matching EMBL:H60047
RHdb:RH78715
dbSTS:STS55456
Identified using the e-PCR software (G. Schuler)"
BASE COUNT 44783 a 52405 c 49132 g 45744 t
ORIGIN

Query Match      0.3%; Score 71; DB 84; Length 192064;
Best Local Similarity 100.0%; Pred. No. 3.9e-27;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17370 ggagtttcacttcttccagcagtgagtgcaatgcggtgatctcagctcagctgcaac 17429
|||||
Db 91171 GGAGTTTCACCTCTTCTGCCAGGCTGGAGTGCATGGCGTGATCTAGCTCACTGCAAC 91230
|||||
QY 17430 ctcgcgcctccc 17440
|||||
Db 91231 CTCGCGCTCCC 91241

RESULT 68
AC087691/c
LOCUS      193816 bp.      DNA      HTG      17-FEB-2001
DEFINITION Homo sapiens chromosome 17 clone RP11-9B11 map 17, WORKING DRAFT
SEQUENCE, 33 unordered pieces.
AC087691
VERSION    AC087691.2 GI:12957885
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 193816)
            Birren,B., Linton,L., Nusbaum,C. and Lander,E.
            Homo sapiens chromosome 17, clone RP11-9B11
            Unpublished
            2 (bases 1 to 193816)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Baran,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Fato,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEvan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Minova,T., Mienga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,

```

Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 17, 2001 this sequence version replaced gi:12229466.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12351
Center clone name: 9_B.11
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 183023 bases at least Q40
Consensus quality: 187299 bases at least Q30
Consensus quality: 189063 bases at least Q20
Insert size: 190616; sum-of-contigs
Quality coverage: 6.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 551: contig of 551 bp in length
* 552 651: gap of 100 bp
* 652 1331: contig of 680 bp in length
* 1332 1431: gap of 100 bp
* 1432 2312: contig of 881 bp in length
* 2313 2412: gap of 100 bp
* 2413 3112: contig of 700 bp in length
* 3113 3212: gap of 100 bp
* 3213 4228: contig of 1016 bp in length
* 4229 4328: gap of 100 bp
* 4329 5482: contig of 1154 bp in length
* 5483 5582: gap of 100 bp
* 5583 7044: contig of 1462 bp in length
* 7045 7144: gap of 100 bp
* 7145 8149: contig of 1005 bp in length
* 8150 8249: gap of 100 bp
* 8250 9789: contig of 1540 bp in length
* 9790 9889: gap of 100 bp
* 9890 11285: contig of 1396 bp in length
* 11286 11385: gap of 100 bp
* 11386 12367: contig of 982 bp in length
* 12368 12467: gap of 100 bp
* 12468 13819: contig of 1352 bp in length
* 13820 13919: gap of 100 bp
* 13920 15542: contig of 1623 bp in length
* 15543 15642: gap of 100 bp
* 15643 17701: contig of 2059 bp in length
* 17702 17801: gap of 100 bp
* 17802 20378: contig of 2577 bp in length
* 20379 20478: gap of 100 bp
* 20479 22935: contig of 2457 bp in length
* 22936 23035: gap of 100 bp
* 23036 25203: contig of 2168 bp in length

* 25204 25303: gap of 100 bp
* 25304 57425: contig of 32122 bp in length
* 57426 57525: gap of 100 bp
* 57526 60236: contig of 2711 bp in length
* 60237 60336: gap of 100 bp
* 60337 62717: contig of 2381 bp in length
* 62718 62817: gap of 100 bp
* 62818 67206: contig of 4389 bp in length
* 67207 67306: gap of 100 bp
* 67307 70551: contig of 3245 bp in length
* 70552 70651: gap of 100 bp
* 70652 73003: contig of 2352 bp in length
* 73004 73103: gap of 100 bp
* 73104 77805: contig of 4702 bp in length
* 77806 77905: gap of 100 bp
* 77906 86152: contig of 8247 bp in length
* 86153 86252: gap of 100 bp
* 86253 96092: contig of 9840 bp in length
* 96093 96192: gap of 100 bp
* 96193 107607: contig of 11415 bp in length
* 107608 107707: gap of 100 bp
* 107708 119748: contig of 12041 bp in length
* 119749 119848: gap of 100 bp
* 119849 129971: contig of 10123 bp in length
* 129972 130071: gap of 100 bp
* 130072 147185: contig of 17114 bp in length
* 147186 147285: gap of 100 bp
* 147286 166313: contig of 19028 bp in length
* 166314 166413: gap of 100 bp
* 166414 191704: contig of 25291 bp in length
* 191705 191804: gap of 100 bp
* 191805 193816: contig of 2012 bp in length.

FEATURES

source

Location/Qualifiers
1..193816
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="Rp11-9B11"
/clone_lib="RPC1-11 Human Male BAC"
1..351
/note="assembly_fragment"
652..1331
/note="assembly_fragment"
1432..2312
/note="assembly_fragment"
2413..3112
/note="assembly_fragment"
3213..4228
/note="assembly_fragment"
4329..5482
/note="assembly_fragment"
5583..7044
/note="assembly_fragment"
7145..8149
/note="assembly_fragment"
8250..9789
/note="assembly_fragment"
9890..11285
/note="assembly_fragment"
11386..12367
/note="assembly_fragment"
12468..13819
/note="assembly_fragment"
13920..15542
/note="assembly_fragment"
15643..17701
/note="assembly_fragment"
17802..20378
/note="assembly_fragment"
20479..22935
/note="assembly_fragment"
23036..25203

misc_feature
/note="assembly_fragment"
25304..57425
misc_feature
/note="assembly_fragment"
57526..60236
misc_feature
/note="assembly_fragment"
60337..62717
misc_feature
/note="assembly_fragment"
62818..67206
misc_feature
/note="assembly_fragment"
67307..70551
misc_feature
/note="assembly_fragment"
70652..73003
misc_feature
/note="assembly_fragment"
73104..77805
misc_feature
/note="assembly_fragment"
77906..86152
misc_feature
/note="assembly_fragment"
86253..96092
misc_feature
/note="assembly_fragment"
96193..107607
misc_feature
/note="assembly_fragment"
107708..119748
misc_feature
/note="assembly_fragment"
119849..129971
/note="assembly_fragment"

Query Match 0.3%; Score 71; DB 77; Length 193816;

Best Local Similarity 100.0%; Pred. No. 3.9e-27; Mismatches 0; Indels 0; Gaps 0;

Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17379 ctctgttgcaggctggagtgcaatggcggtgatctcagctcactgcaacctccgcctc 17438

|||||

Db 5351 CTCTTGTGCCAGGCTGGAGTGCATGGGTGATCTCAGCTCAGCTCGACCTCGCGCTC 5292

QY 17439 cggggttcgaag 17449

|||||

Db 5291 CCGGGTTCAAG 5281

RESULT 69

AC005291 198582 bp DNA PRI
LOCUS Homo sapiens chromosome 17, clone hRPK.401_O_9, complete sequence.
DEFINITION AC005291
ACCESSION AC005291
VERSION AC005291.1 GI:3402737
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 198582)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hRPK.401_O_9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 198582)
AUTHORS Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depeyre,E.,
Devon,K., Dewar,K., Doneelan,L., Etemadi,S., Ferreira,P.,
FitzHugh,W., Forrest,C., Funke,C., Gage,D., Gage,D., Gardyna,S.,
Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
Strange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
Zody,M.
Direct Submission
Submitted (17-JUL-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
TITLE
JOURNAL

REFERENCE
AUTHORS

3 (bases 1 to 198582)
 Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
 Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J.,
 Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M.,
 Collins,S., Collamore,A., Cooke,P., Corliss,D., Depayre,E.,
 Devon,K., Dewar,K., Donelan,L., Ferrelira,P., FitzHugh,W.,
 Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraigery,K., Grant,G.,
 Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C.,
 Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczyk,J.,
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
 Nahr,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,
 Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
 Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
 Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
 Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
 Ye,W.J., Zhao,J. and Zody,M.

TITLE
JOURNAL

Submitted (07-AUG-1998) Whitehead Institute/MIT Center for Genome

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 7, 1998 this sequence version replaced gi:3399679.
 All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)

FEATURES

Location/Qualifiers

source

1..198582
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="hRPK.401_O_9"
 /clone_lib="RPCI-II human BAC library"
 /map="17"
 /chromosome="17"

repeat_region	59..160	/rpt_family="L2"
repeat_region	162..429	/rpt_family="L1MC5"
repeat_region	431..578	/rpt_family="AluSg"
repeat_region	580..1122	complement(578..1122)
repeat_region	1123..1427	/rpt_family="MER41B"
repeat_region	1428..1523	/rpt_family="AluY"
repeat_region	1524..2090	complement(1428..1523)
repeat_region	2091..2402	/rpt_family="MER41B"
repeat_region	2403..2414	/rpt_family="AluSx"
repeat_region	2415..2715	/rpt_family="AluJo"
repeat_region	2716..3021	complement(2843..2927)
repeat_region	3022..3315	/rpt_family="MER57-internal"
repeat_region	3316..4012	/rpt_family="AluSp"
repeat_region	4013..4623	complement(4012..4318)
repeat_region	4624..5032	/rpt_family="AluSx"
repeat_region	5033..5095	complement(4485..4623)
repeat_region	5096..5678	/rpt_family="MER5B"
repeat_region	5679..6440	complement(4726..5032)
repeat_region	6441..6559	/rpt_family="AluSx"
repeat_region	6560..6866	complement(5033..5095)
repeat_region	6867..7224	/rpt_family="MER5B"
repeat_region	7225..7386	complement(5678..5998)
repeat_region	7387..7515	/rpt_family="AluSx"
repeat_region	7516..7543	complement(6440..6559)
repeat_region	7544..7624	/rpt_family="FLAM_A"
repeat_region	7625..7966	complement(6564..6866)
repeat_region	7967..8112	/rpt_family="AluSc"
repeat_region	8113..8247	complement(6872..7224)
repeat_region	8248..8406	/rpt_family="MLT1F"
repeat_region	8407..8515	complement(7228..7386)
repeat_region	8516..8589	/rpt_family="AluSp"
repeat_region	8590..8624	complement(7409..7515)
repeat_region	8625..8626	/rpt_family="MLT1F"
repeat_region	8627..8628	/rpt_family="MER91A"
repeat_region	8629..8630	complement(7677..7966)
repeat_region	8631..8632	
repeat_region	8633..8634	
repeat_region	8635..8636	
repeat_region	8637..8638	
repeat_region	8639..8640	
repeat_region	8641..8642	
repeat_region	8643..8644	
repeat_region	8645..8646	
repeat_region	8647..8648	
repeat_region	8649..8650	
repeat_region	8651..8652	
repeat_region	8653..8654	
repeat_region	8655..8656	
repeat_region	8657..8658	
repeat_region	8659..8660	
repeat_region	8661..8662	
repeat_region	8663..8664	
repeat_region	8665..8666	
repeat_region	8667..8668	
repeat_region	8669..8670	
repeat_region	8671..8672	
repeat_region	8673..8674	
repeat_region	8675..8676	
repeat_region	8677..8678	
repeat_region	8679..8680	
repeat_region	8681..8682	
repeat_region	8683..8684	
repeat_region	8685..8686	
repeat_region	8687..8688	
repeat_region	8689..8690	
repeat_region	8691..8692	
repeat_region	8693..8694	
repeat_region	8695..8696	
repeat_region	8697..8698	
repeat_region	8699..8700	
repeat_region	8701..8702	
repeat_region	8703..8704	
repeat_region	8705..8706	
repeat_region	8707..8708	
repeat_region	8709..8710	
repeat_region	8711..8712	
repeat_region	8713..8714	
repeat_region	8715..8716	
repeat_region	8717..8718	
repeat_region	8719..8720	
repeat_region	8721..8722	
repeat_region	8723..8724	
repeat_region	8725..8726	
repeat_region	8727..8728	
repeat_region	8729..8730	
repeat_region	8731..8732	
repeat_region	8733..8734	
repeat_region	8735..8736	
repeat_region	8737..8738	
repeat_region	8739..8740	
repeat_region	8741..8742	
repeat_region	8743..8744	
repeat_region	8745..8746	
repeat_region	8747..8748	
repeat_region	8749..8750	
repeat_region	8751..8752	
repeat_region	8753..8754	
repeat_region	8755..8756	
repeat_region	8757..8758	
repeat_region	8759..8760	
repeat_region	8761..8762	
repeat_region	8763..8764	
repeat_region	8765..8766	
repeat_region	8767..8768	
repeat_region	8769..8770	
repeat_region	8771..8772	
repeat_region	8773..8774	
repeat_region	8775..8776	
repeat_region	8777..8778	
repeat_region	8779..8780	
repeat_region	8781..8782	
repeat_region	8783..8784	
repeat_region	8785..8786	
repeat_region	8787..8788	
repeat_region	8789..8790	
repeat_region	8791..8792	
repeat_region	8793..8794	
repeat_region	8795..8796	
repeat_region	8797..8798	
repeat_region	8799..8800	
repeat_region	8801..8802	
repeat_region	8803..8804	
repeat_region	8805..8806	
repeat_region	8807..8808	
repeat_region	8809..8810	
repeat_region	8811..8812	
repeat_region	8813..8814	
repeat_region	8815..8816	
repeat_region	8817..8818	
repeat_region	8819..8820	
repeat_region	8821..8822	
repeat_region	8823..8824	
repeat_region	8825..8826	
repeat_region	8827..8828	
repeat_region	8829..8830	
repeat_region	8831..8832	
repeat_region	8833..8834	
repeat_region	8835..8836	
repeat_region	8837..8838	
repeat_region	8839..8840	
repeat_region	8841..8842	
repeat_region	8843..8844	
repeat_region	8845..8846	
repeat_region	8847..8848	
repeat_region	8849..8850	
repeat_region	8851..8852	
repeat_region	8853..8854	
repeat_region	8855..8856	
repeat_region	8857..8858	
repeat_region	8859..8860	
repeat_region	8861..8862	
repeat_region	8863..8864	
repeat_region	8865..8866	
repeat_region	8867..8868	
repeat_region	8869..8870	
repeat_region	8871..8872	
repeat_region	8873..8874	
repeat_region	8875..8876	
repeat_region	8877..8878	
repeat_region	8879..8880	
repeat_region	8881..8882	
repeat_region	8883..8884	
repeat_region	8885..8886	
repeat_region	8887..8888	
repeat_region	8889..8890	
repeat_region	8891..8892	
repeat_region	8893..8894	
repeat_region	8895..8896	
repeat_region	8897..8898	
repeat_region	8899..8900	
repeat_region	8901..8902	
repeat_region	8903..8904	
repeat_region	8905..8906	
repeat_region	8907..8908	
repeat_region	8909..8910	
repeat_region	8911..8912	
repeat_region	8913..8914	
repeat_region	8915..8916	
repeat_region	8917..8918	
repeat_region	8919..8920	
repeat_region	8921..8922	
repeat_region	8923..8924	
repeat_region	8925..8926	
repeat_region	8927..8928	
repeat_region	8929..8930	
repeat_region	8931..8932	
repeat_region	8933..8934	
repeat_region	8935..8936	
repeat_region	8937..8938	
repeat_region	8939..8940	
repeat_region	8941..8942	
repeat_region	8943..8944	
repeat_region	8945..8946	
repeat_region	8947..8948	
repeat_region	8949..8950	
repeat_region	8951..8952	
repeat_region	8953..8954	
repeat_region	8955..8956	
repeat_region	8957..8958	
repeat_region	8959..8960	
repeat_region	8961..8962	
repeat_region	8963..8964	
repeat_region	8965..8966	
repeat_region	8967..8968	
repeat_region	8969..8970	
repeat_region	8971..8972	
repeat_region	8973..8974	
repeat_region	8975..8976	
repeat_region	8977..8978	
repeat_region	8979..8980	
repeat_region	8981..8982	
repeat_region	8983..8984	
repeat_region	8985..8986	
repeat_region	8987..8988	
repeat_region	8989..8990	
repeat_region	8991..8992	
repeat_region	8993..8994	
repeat_region	8995..8996	
repeat_region	8997..8998	
repeat_region	8999..9000	
repeat_region	9001..9002	
repeat_region	9003..9004	
repeat_region	9005..9006	
repeat_region	9007..9008	
repeat_region	9009..9010	
repeat_region	9011..9012	
repeat_region	9013..9014	
repeat_region	9015..9016	
repeat_region	9017..9018	
repeat_region	9019..9020	
repeat_region	9021..9022	
repeat_region	9023..9024	
repeat_region	9025..9026	
repeat_region	9027..9028	
repeat_region	9029..9030	
repeat_region	9031..9032	
repeat_region	9033..9034	
repeat_region	9035..9036	
repeat_region	9037..9038	
repeat_region	9039..9040	
repeat_region	9041..9042	
repeat_region	9043..9044	
repeat_region	9045..9046	
repeat_region	9047..9048	
repeat_region	9049..9050	
repeat_region	9051..9052	
repeat_region	9053..9054	
repeat_region	9055..9056	
repeat_region	9057..9058	
repeat_region	9059..9060	
repeat_region	9061..9062	
repeat_region	9063..9064	
repeat_region	9065..9066	
repeat_region	9067..9068	
repeat_region	9069..9070	
repeat_region	9071..9072	
repeat_region	9073..9074	
repeat_region	9075..9076	
repeat_region	9077..9078	
repeat_region	9079..9080	
repeat_region	9081..9082	
repeat_region	9083..9084	
repeat_region	9085..9086	
repeat_region	9087..9088	
repeat_region	9089..9090	
repeat_region	9091..9092	
repeat_region	9093..9094	
repeat_region	9095..9096	
repeat_region	9097..9098	
repeat_region	9099..9100	
repeat_region	9101..9102	
repeat_region	9103..9104	
repeat_region	9105..9106	
repeat_region	9107..9108	
repeat_region	9109..9110	
repeat_region	9111..9112	
repeat_region	9113..9114	
repeat_region	9115..9116	
repeat_region	9117..9118	
repeat_region	9119..9120	
repeat_region	9121..9122	
repeat_region	9123..9124	
repeat_region	9125..9126	
repeat_region	9127..9128	
repeat_region	9129..9130	
repeat_region	9131..9132	
repeat_region	9133..9134	
repeat_region	9135..9136	
repeat_region	9137..9138	
repeat_region	9139..9140	
repeat_region	9141..9142	
repeat_region	9143..9144	
repeat_region	9145..9146	
repeat_region	9147..9148	
repeat_region	9149..9150	
repeat_region	9151..9152	
repeat_region	9153..9154	
repeat_region	9155..9156	
repeat_region	9157..9158	
repeat_region	9159..9160	
repeat_region	9161..9162	
repeat_region	9163..9164	
repeat_region	9165..9166	
repeat_region	9167..9168	
repeat_region	9169..9170	
repeat_region	9171..9172	
repeat_region	9173..9174	
repeat_region	9175..9176	
repeat_region	9177..9178	
repeat_region	9179..9180	
repeat_region	9181..9182	

```
repeat_region complement(22333..22557)
repeat_region /rpt_family="MLT1J"
repeat_region complement(22715..22824)
repeat_region /rpt_family="MIR"
repeat_region complement(22825..23109)
repeat_region /rpt_family="AluSp"
repeat_region complement(23110..23192)
repeat_region /rpt_family="MIR"
repeat_region 24077..24285
repeat_region /rpt_family="AluY"
repeat_region complement(24558..24863)
repeat_region /rpt_family="AluSx"
repeat_region complement(25044..25171)
repeat_region /rpt_family="MLT1A1"
repeat_region 25172..25469
repeat_region /rpt_family="AluSq"
repeat_region complement(25470..25682)
repeat_region /rpt_family="MLT1A1"
repeat_region complement(25893..26185)
repeat_region /rpt_family="MLT1E"
repeat_region 26194..26267
repeat_region /rpt_family="purine-rich"

Query Match 0.3%; Score 71; DB 85; Length 198582;
Best Local Similarity 100.0%; Pred. No. 3.9e-27;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17379 cttctgttgcacagctggagtcgaatggcgatcagtcactgcaacctcgcctc 17438
|||||
Db 7254 CTCGTGTGGCCAGCTGGAGTGAATGGCGTATCTCAGTCACTGCAACCTCCGCCTC 7313
|||||
QY 17439 ccgggttcaag 17449
|||||
Db 7314 CCGGTTCAAG 7324

RESULT 70
AC069250/c
LOCUS AC069250 203748 bp DNA HTG 12-APR-2001
DEFINITION Homo sapiens chromosome 17 clone RP11-9B11, WORKING DRAFT SEQUENCE,
16 unordered pieces.
ACCESSION AC069250
VERSION AC069250.14 GI:13605982
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 203748)
AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Fedorpiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
Unpublished
JOURNAL 2 (bases 1 to 203748)
REFERENCE Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
AUTHORS Fedorpiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
Direct Submission
TITLE Submitted (22-MAY-2000) DNA Sequencing and Technology Center,
JOURNAL Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
COMMENT On Apr 12, 2001 this sequence version replaced gi:9828651.
----- Genome Center
Center: Stanford DNA Sequencing and Technology Development
Center
Center code: SDSTDC
Web site: http://sequence-www.stanford.edu/group/human/
Contact: hum-info@sequence.stanford.edu
----- Project Information
Center project name: 941
```

```
Center clone name: RP11-9B11
----- Summary Statistics
Sequencing Vector: M13mp18; X02513
Chemistry: Dye-primer; 26% of reads
Assembly: Dye-terminator Big Dye; 74% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 195554 bases at least Q40
Consensus quality: 197606 bases at least Q30
Consensus quality: 198578 bases at least Q20
Insert size: 197955; agarose-fp
Insert size: 202248; sum-of-contigs
Quality coverage: 11.1x in Q20 bases; agarose-fp
Quality coverage: 10.9x in Q20 bases; sum-of-contigs.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1678: contig of 1678 bp in length
* 1679: gap of unknown length
* 1779: contig of 3085 bp in length
* 4864: gap of unknown length
* 4964: contig of 4602 bp in length
* 9565: gap of unknown length
* 9566: contig of 5441 bp in length
* 15106: gap of unknown length
* 15107: contig of 9668 bp in length
* 15207: gap of unknown length
* 24875: contig of 7148 bp in length
* 24975: gap of unknown length
* 32123: contig of 10007 bp in length
* 32223: gap of unknown length
* 42229: contig of 9857 bp in length
* 52186: gap of unknown length
* 52187: contig of 11637 bp in length
* 52286: gap of unknown length
* 63924: gap of unknown length
* 64024: contig of 11991 bp in length
* 76015: gap of unknown length
* 76115: contig of 12150 bp in length
* 88265: gap of unknown length
* 88365: contig of 13318 bp in length
* 101683: gap of unknown length
* 101783: contig of 14673 bp in length
* 116456: gap of unknown length
* 116556: contig of 17128 bp in length
* 133684: gap of unknown length
* 133784: contig of 29910 bp in length
* 163694: gap of unknown length
* 163794: contig of 39955 bp in length.
FEATURES
Location/Qualifiers
source 1..203748
organism="Homo sapiens"
db_xref="taxon:9606"
chromosome="17"
clone="RP11-9B11"
clone_lib="RPC1 human BAC library 11"
1..1678
note="assembly_name:Contig33"
misc_feature 1779..4863
note="assembly_name:Contig35"
misc_feature 4964..9565
note="assembly_name:Contig36"
misc_feature 9666..15106
note="assembly_name:Contig37"
misc_feature 15207..24874
note="assembly_name:Contig38"
misc_feature 24975..32122
note="assembly_name:Contig39"
misc_feature 32223..42229
note="assembly_name:Contig40"
```

misc_feature clone_end:SP6"
 42330..52186
 /note="assembly_name:Contig41"
 misc_feature 52287..63923
 /note="assembly_name:Contig42"
 64034..76014
 /note="assembly_name:Contig43"
 misc_feature 76115..88264
 /note="assembly_name:Contig44"
 88365..101682
 /note="assembly_name:Contig45"
 misc_feature 101783..116455
 /note="assembly_name:Contig46"
 misc_feature 116556..133683
 /note="assembly_name:Contig47"
 misc_feature 133784..163693
 /note="assembly_name:Contig48"
 misc_feature 163794..203748
 /note="assembly_name:Contig49"
 BASE COUNT 52678 a 48865 c 47969 g 52716 t 1520 others
 ORIGIN

Query Match 0.3% Score 71; DB 74; Length 203748;
 Best Local Similarity 100.0%; Pred. No. 3.9e-27;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 17379 ctctgtgtgccagctgagtgcaatggctgctcagctcactgcaacctgcctc 17438
 Db 39384 CTCCTGTGTGCCAGGCTGAGTGCATGGCGTATCTAGCTCACTGCAACCTCGCGCTC 39325
 QY 17439 cgggttccaag 17449
 Db 39324 CCGGTTTCAAG 39314

RESULT 71
 AC007639/c 94205 bp DNA PRI 19-JUN-1999
 LOCUS Homo sapiens chromosome 17, clone hRPK.57_A_1, complete sequence.
 DEFINITION AC007639
 VERSION AC007639.5 GI:5103882
 KEYWORDS HTG.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 94205)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 17, clone hRPK.57_A_1
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 94205)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
 Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
 Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
 Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
 Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
 Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
 Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
 Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
 Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
 Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
 Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo.A., Wagner,A.,
 Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
 Direct Submission

TITLE
 JOURNAL
 Submitted (22-MAY-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE 3 (bases 1 to 94205)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,

Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
 Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
 Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
 Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
 Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
 Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
 Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
 Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
 Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
 Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo.A., Wagner,A.,
 Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (19-JUN-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT
 On Jun 19, 1999 this sequence version replaced gi:5091642.
 All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Only 94205 base pairs from the middle of this clone are being
 submitted. The remainder overlaps either accession number AC006948
 (WICGR project L564) or accession number AC005243 (WICGR project
 L343).

FEATURES	Location/Qualifiers
source	1..94205
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="hrPK.57_A_1"
	/clone_lib="RPCI-11 human BAC library"
	/map="17"
	/chromosome="17"
repeat_region	complement(46..419)
	/rpt_family="L1MC4"
repeat_region	420..828
	/rpt_family="L1PA8"
repeat_region	complement(829..1560)
	/rpt_family="L1MC4"
repeat_region	1992..2248
	/rpt_family="MER102"
unsure	3579..3593
	/note="Single-stranded coverage."
repeat_region	4578..4604
	/rpt_family="(T)n"
repeat_region	4748..4790
	/rpt_family="(TTTA)n"
repeat_region	6378..6451
	/rpt_family="L2"
repeat_region	6855..6888
	/rpt_family="AT-rich"
repeat_region	6895..7210
	/rpt_family="AluSg"
repeat_region	complement(7212..7350)
	/rpt_family="AluSg"
repeat_region	7969..8003
	/rpt_family="(CA)n"
repeat_region	complement(8269..8580)
	/rpt_family="AluSg"
repeat_region	complement(10135..10187)
	/rpt_family="MIR"
repeat_region	10387..10528
	/rpt_family="L2"
repeat_region	complement(11180..11214)
	/rpt_family="MIR"
repeat_region	complement(11215..11566)
	/rpt_family="THE1A"
repeat_region	complement(11567..11746)
	/rpt_family="MIR"
repeat_region	complement(13420..13729)
	/rpt_family="AluSg"
repeat_region	complement(13979..14209)
	/rpt_family="L2"

```
repeat_region 14248..14333 /rpt_family="MIR"
repeat_region complement(14378..14505)
repeat_region /rpt_family="L2"
repeat_region complement(14550..14709)
repeat_region /rpt_family="L2"
repeat_region 14732..15029 /rpt_family="AluJo"
repeat_region 15064..15107 /rpt_family="GGAA)n"
repeat_region 16502..16535 /rpt_family="AT-rich"
repeat_region complement(16619..16676)
repeat_region /rpt_family="MIR"
repeat_region 16687..16987
repeat_region /rpt_family="AluSq"
repeat_region 17547..17601 /rpt_family="(TA)n"
repeat_region 18214..18251 /rpt_family="GA-rich"
repeat_region 19150..19211 /rpt_family="ORSL"
repeat_region 19212..19506 /rpt_family="AluSq"
repeat_region 19507..19657 /rpt_family="ORSL"
repeat_region 20843..20934 /rpt_family="MER5A"
repeat_region 20959..21145 /rpt_family="MER5B"
repeat_region complement(22241..22322)
repeat_region /rpt_family="MIR"
repeat_region complement(23032..23219)
repeat_region /rpt_family="MIR"
repeat_region 23741..23768 /rpt_family="(GAAAA)n"
repeat_region 23983..24213 /rpt_family="MIR"
repeat_region 24241..24350 /rpt_family="MIR"
repeat_region complement(25435..25547)
repeat_region /rpt_family="L2"
repeat_region complement(25548..25894)
repeat_region /rpt_family="THE1B"
repeat_region complement(25991..26310)
repeat_region /rpt_family="L2"
repeat_region 26735..26768 /rpt_family="(A)n"
repeat_region 27776..27796 /rpt_family="(TG)n"
repeat_region 28780..28818 /rpt_family="AT-rich"
repeat_region 30798..30907 /rpt_family="MIR"
repeat_region complement(31413..31530)
repeat_region /rpt_family="FLAM-A"
repeat_region complement(31749..31810)
repeat_region /rpt_family="L2"
repeat_region 32215..32498 /rpt_family="AluJo"
repeat_region 32510..32936 /rpt_family="MLT1C"
repeat_region complement(33616..34059)
repeat_region /rpt_family="MLT1D"
repeat_region 34840..34865 /rpt_family="(GA)n"
repeat_region 36201..36507 /rpt_family="AluSc"
repeat_region complement(36512..36663)
repeat_region /rpt_family="MIR"
repeat_region 36788..36826 /rpt_family="(TTCA)n"
repeat_region 38452..38773

Query Match 0.3%; Score 70; DB 86; Length 94205;
Best Local Similarity 100.0%; Pred. No. 1.3e-26; Mismatches 0; Indels 0; Gaps 0;
Matches 70; Conservative 0;

QY 4718 ttgtatttttagtagacggggtttccaccatgttgccaggatagctgcgtctctga 4777
|||||
Db 89818 TTGTATTTTGTAGTAGAGACGGGGTTTCACCATGTTGCCAGGATAGTCTCGATCTCTGA 89759

QY 4778 ccttgatc 4787
|||||
Db 89758 CCTTGTGATC 89749

RESULT 72
AL133282
LOCUS AL133282 130526 bp DNA PRI 26-MAY-2000
DEFINITION Human DNA sequence from clone RP11-264C15 on chromosome 9q32-34.11,
complete sequence.
ACCESSION AL133282
VERSION AL133282.15 GI:8246854
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 130526)
AUTHORS Sehra,H.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
On Jun 4, 2000 this sequence version replaced gi:8218296.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
```

RP11-264C15 is from the library RPCI-11.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR: pRACE3.6>

IMPORTANT: This sequence is not the entire insert of clone RP11-264C15. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-264C15 is at 1 in this sequence. The true left end of clone RP11-67K19 is at 130427 in this sequence. The true right end of clone RP11-45A16 is at 784 in this sequence.

FEATURES

source
Location/Qualifiers
1..130526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="q32-q34.11"
/clone="RP11-264C15"
/clone_lib="RPCI-11.1"
38991 a 27088 c 27582 g 36865 t

BASE COUNT
ORIGIN

Query Match 0.3%; Score 70; DB 89; Length 130526;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10343 tttttagacagatctgtctgtccaccagctgagtcagtgctgctatcttgctca 10402
|||||
Db 31411 TTTTGGACAGAGTCGCTCTGTCCACGAGCTGGAGTCGATCTTGCTCA 31470

QY 10403 ctgcacacctc 10412
|||||
Db 31471 CTGCACCTC 31480

RESULT 73
AC026144/c

LOCUS AC026144 147990 bp DNA HTG 05-APR-2000
Homo sapiens chromosome 4 clone RP11-106H1 map 4, WORKING DRAFT
SEQUENCE, 17 unordered pieces.

ACCESSION AC026144
VERSION AC026144.2 GI:7417853
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 147990)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 4, clone RP11-106H1
Unpublished
2 (bases 1 to 147990)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campolano,A., Castle,A., Choemel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivat,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J.,

TITLE

JOURNAL

COMMENT

Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 5, 2000 this sequence version replaced gi:7264214.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8392

Center clone name: 106_H1

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of pieces

Assembly program: Phrap; version 0.960731

Consensus quality: 137712 bases at least Q40

Consensus quality: 142871 bases at least Q30

Consensus quality: 144908 bases at least Q20

Insert size: 153000; agarose-fp

Insert size: 146390; sum-of-contigs

Quality coverage: 4.2 in Q20 bases; agarose-fp

Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1059: contig of 1059 bp in length
* 1060 1159: gap of 100 bp
* 1160 2331: contig of 1172 bp in length
* 2332 2431: gap of 100 bp
* 2432 4117: contig of 1686 bp in length
* 4118 4217: gap of 100 bp
* 4218 6670: contig of 2453 bp in length
* 6671 6770: gap of 100 bp
* 6771 9681: contig of 2911 bp in length
* 9682 9781: gap of 100 bp
* 9782 14364: contig of 4783 bp in length
* 14365 14664: gap of 100 bp
* 14665 22015: contig of 7351 bp in length
* 22016 22115: gap of 100 bp
* 22116 29642: contig of 7527 bp in length
* 29643 29742: gap of 100 bp
* 29743 37004: contig of 7262 bp in length
* 37005 37104: gap of 100 bp
* 37105 45548: contig of 8444 bp in length
* 45549 45648: gap of 100 bp
* 45649 54452: contig of 8804 bp in length
* 54453 54552: gap of 100 bp
* 54553 65398: contig of 10846 bp in length
* 65399 65498: gap of 100 bp
* 65499 78031: contig of 12533 bp in length
* 78032 78131: gap of 100 bp
* 78132 94963: contig of 16832 bp in length
* 94964 95063: gap of 100 bp
* 95064 111774: contig of 16711 bp in length
* 111775 111874: gap of 100 bp
* 111875 127997: contig of 16123 bp in length
* 127998 128097: gap of 100 bp
* 128098 147990: contig of 19893 bp in length.

FEATURES

Location/Qualifiers
1..147990

/organism="Homo sapiens"
/db_xref="taxon:9606"

source

```
/chromosome="4"
/map="4"
/clone="RP11-106H1"
/clone_lib="RPC1-11 Human Male BAC"
1. 1059
/note="assembly_fragment"
1160. 2331
/note="assembly_fragment"
2432. 4117
/note="assembly_fragment"
4218. 6670
/note="assembly_fragment"
6771. 9681
/note="assembly_fragment
clone_end:T7
vector_side:right"
9782. 14564
/note="assembly_fragment"
14665. 22015
/note="assembly_fragment"
22116. 29642
/note="assembly_fragment"
29743. 37004
/note="assembly_fragment"
37105. 45548
/note="assembly_fragment"
45649. 54452
/note="assembly_fragment
clone_end:SP6
vector_side:right"
54553. 65398
/note="assembly_fragment"
65499. 78031
/note="assembly_fragment"
78132. 94963
/note="assembly_fragment"
95064. 111174
/note="assembly_fragment"
111875. 127997
/note="assembly_fragment"
128098. 147990
/note="assembly_fragment"
BASE COUNT 44391 a 28973 c 29088 g 43938 t 1600 others
ORIGIN

Query Match 0.3%; Score 70; DB 69; Length 147990;
Best Local Similarity 100.0%; Pred.No.1.4e-26;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17556 agcgtgctcgaactctgacctcagtgatccaccacctcagctcccaagtgttg 17615
|||||
Db 60891 AGCGTGTCTCGAACTCTGACCTCAGGTGATCCACCCAGCTCCCAAGTGTG 60832
|||||

QY 17616 ggattacagg 17625
|||||
Db 60831 GGATTACAGG 60822

RESULT 74
CNS01DUW/c
LOCUS
DEFINITION Homo sapiens chromosome 14 clone R-219E7, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL133371
VERSION AL133371.3 GI:11611141
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 162472)
AUTHORS Genoscope. "
```

Direct Submission
Submitted (30-JAN-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On Dec 9, 2000 this sequence version replaced gi:6634066.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr

IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc. . even if efforts are made to eliminate these
contaminating sequences. The following BAC sequence is oriented
from the T7 to the SP6 end.
Downstream BAC (overlapping the T7 end) : R-903H12
Upstream BAC (overlapping the SP6 end) : R-84C10 (AC-AL355922)
----- Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 8.64x in Q20 bases; sum-of-contigs

Overall quality chart :
Range : bases
0 :
1 - 9 :
10 - 19 :
20 - 29 : 1
30 - 39 : 10
40 - 49 : 513
50 - 59 : 1389
60 - 69 : 4390
70 - 79 : 15474
80 - 89 : 48563
90 - 99 : 92132

Percentage of bases with a quality value >= 40 : 99 %.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
Location/Qualifiers
1..162472
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-219E7"
/clone_lib="RPC1-11"
66..342
/note="matching EMBL:X76383
RHdb:RH70907
dbSTS:STS50761
Identified using the e-PCR software (G. Schuler)"
34326..34475
/note="matching EMBL:G15182
RHdb:RH14231
dbSTS:STS2174
Identified using the e-PCR software (G. Schuler)"
53428..53648
/note="matching EMBL:G14582
RHdb:RH7933
dbSTS:STS1734
Identified using the e-PCR software (G. Schuler)"
144244..144384
/note="matching EMBL:G15094
RHdb:RH29940
dbSTS:STS31745
Identified using the e-PCR software (G. Schuler)"
144244..144399

TITLE
JOURNAL
COMMENT

```
/note="matching EMBL:G11174
RHdb:RH1872
dbSTS:STS25650
Identified using the e-PCR software (G. Schuler)"
/54904....155170
/note="matching EMBL:Z16878
RHdb:RH53522
RHdb:RH42199
RHdb:RH3672
dbSTS:STS28835
Identified using the e-PCR software (G. Schuler)"
/54968...155213
/note="matching EMBL:Z16878
RHdb:RH73593
RHdb:RH49026
RHdb:RH12971
RHdb:RH643
dbSTS:STS4033
Identified using the e-PCR software (G. Schuler)"
BASE COUNT 4788 a 34301 c 35618 g 44655 t
ORIGIN

Query Match      0.3%; Score 70; DB 83; Length 162472;
Best Local Similarity 100.0%; Pred. No. 1.4e-26;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8913 caatctggtcactgcagctccacccacggttcaggatctctctgcctcagcct 8972
|||||
Db 40669 CAATCTCGCTCAGTCGACGCTCCACCTCCACGTTCAAGTGATTCTCTGCTCAGCCT 40610
|||||
QY 8973 cccaagtagc 8982
|||||
Db 40609 CCCAGTAGC 40600

RESULT 75
AC022076
LOCUS AC022076.11 GI:12039055
DEFINITION Homo sapiens chromosome 3 clone RP11-34L7, WORKING DRAFT SEQUENCE,
15 unordered pieces.
ACCESSION AC022076
VERSION AC022076.11
KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 165464)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Albrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieval,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,B., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
```

```
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
Ogih,M., Okunolu,G., Oragunye,N., Oviedo,R., Pacer,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williams,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 165464)
Worley,K.C.
Direct Submission
Submitted (26-JAN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 5, 2001 this sequence version replaced gi:8699653.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HABR
Center clone name: RP11-34L7
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 153622 bases at least Q40
Consensus quality: 159659 bases at least Q30
Consensus quality: 161816 bases at least Q20
Estimated insert size: 160303; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 4.8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 61275: contig of 61275 bp in length
* 61276 61375: gap of unknown length
* 61376 85217: contig of 23842 bp in length
* 85218 85317: gap of unknown length
* 85318 102804: contig of 17487 bp in length
* 102805 102904: gap of unknown length
* 102905 116420: contig of 13516 bp in length
* 116421 116520: gap of unknown length
* 116521 127805: contig of 11285 bp in length
* 127806 127905: gap of unknown length
* 127906 138541: contig of 10636 bp in length
* 138542 138641: gap of unknown length
* 138642 144474: contig of 5833 bp in length
* 144475 144574: gap of unknown length
* 144575 149556: contig of 5082 bp in length
* 149557 149577: gap of unknown length
* 149578 153536: contig of 3780 bp in length
* 153537 153636: gap of unknown length
* 153637 157572: contig of 3936 bp in length
* 157573 157672: gap of unknown length
```

```
* 157673 159023: contig of 1351 bp in length
* 159024 159123: gap of unknown length
* 159124 161694: contig of 2571 bp in length
* 161695 161794: gap of unknown length
* 161795 162817: contig of 1023 bp in length
* 162818 162917: gap of unknown length
* 162918 164327: contig of 1410 bp in length
* 164328 164427: gap of unknown length
* 164428 165464: contig of 1037 bp in length.
FEATURES
    Location/Qualifiers
        1..165464
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosomes="3"
            /clone="RP11-34L7"
BASE COUNT 46537 a 36767 c 36115 g 44624 t 1421 others
ORIGIN
Query Match      0.3%; Score 70; DB 67; Length 165464;
Best Local Similarity 100.0%; Pred No. 1.4e-26;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17555 caggctggtctgaactcctgaactcaggtgacccacccacccagctcccaagtggtt 17614
      |||
Db 2910 CAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCAGCTCCCAAGTGTT 2969
QY 17615 gggattacag 17624
      |||
Db 2970 GGGATTACAG 2979
RESULT 76
AC013244
LOCUS
DEFINITION
AC013244
AC013244
AC013244.32 GI:13654318
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Blincke,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Duggan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
```

```
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,R., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,V., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 169533)
Worley,K.C.
Direct Submission
Submitted (04-NOV-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 17, 2001 this sequence version replaced gi:13626074.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMHT
Center clone name: RP11-60E8
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 163682 bases at least Q40
Consensus quality: 169241 bases at least Q30
Consensus quality: 171788 bases at least Q20
Estimated insert size: 166673; sum-of-contigs estimation
Quality coverage: 8.6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 70967: contig of 70967 bp in length
* 70968 71067: gap of unknown length
* 121613 121613: contig of 50546 bp in length
* 121614 121713: gap of unknown length
* 121714 133088: contig of 11375 bp in length
* 133089 133188: gap of unknown length
* 133189 142835: contig of 9647 bp in length
* 142836 142935: gap of unknown length
* 142936 148941: contig of 6006 bp in length
* 148942 149041: gap of unknown length
* 149042 153535: contig of 4494 bp in length
* 153536 153635: gap of unknown length
* 153636 158767: contig of 5132 bp in length
* 158768 158867: gap of unknown length
* 158868 161357: contig of 2490 bp in length
* 161358 161457: gap of unknown length
* 161458 164237: contig of 2780 bp in length
* 164238 164337: gap of unknown length
* 164338 166237: contig of 1900 bp in length
* 166238 166337: gap of unknown length
* 166338 167987: contig of 1650 bp in length
* 167988 168087: gap of unknown length
* 168088 169533: contig of 1446 bp in length.
Location/Qualifiers
```


source

1..169533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-60E8"

BASE COUNT 49797 a 36145 c 35091 g 47388 t 1112 others
ORIGIN

Query Match 0.3%; Score 70; DB 63; Length 169533;
Best Local Similarity 100.0%; Pred. No. 1.4e-26;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17379 ctctgttccaggctgagtgcaatggcgtgacatcagctcaactgcacctccgcctc 17438

|||||
Db 30763 CTTCTGTTCCCGAGGTGGATGCGTGAATCTGAGTCACAGCACTCCGCCCTC 30822

QY 17439 ccgggttcaa 17448

|||||

Db 30823 CCGGGTTCAA 30832

RESULT 77

AC007940/c

LOCUS

DEFINITION Homo sapiens clone RP11-44C14, WORKING DRAFT SEQUENCE, 11 unordered
pieces.

AC007940

VERSION AC007940.3 GI:8072418

KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 171841)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens, clone RP11-44C14

Unpublished

2 (bases 1 to 171841)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,

Castle,A., Cerny,J., Collangelo,M., Collins,S., Collamore,A.,

Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,

Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,

Funk,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,

Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,

Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,

Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,

Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,

Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,

Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,

Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,

Thesaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,

Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

Direct Submission

Submitted (01-JUL-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 25, 2000 this sequence version replaced gi:7321478.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1050

Center clone name: 44_C14

----- Summary Statistics

Sequencing vector: M13; M77815; 88% of reads

Sequencing vector: Plasmid; n/a; %0.f% of reads

11.8734793187348Chemistry: Dye-primer-amersham; 88% of reads

Chemistry: Dye-terminator Big Dye; 12% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 163415 bases at least Q40
Consensus quality: 167585 bases at least Q30
Consensus quality: 169143 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 170841; sum-of-contigs
Quality covera.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 1291 1390: gap of 100 bp in length
* 1391 2428: contig of 1038 bp in length
* 2429 2528: gap of 100 bp
* 2529 4409: contig of 1881 bp in length
* 4410 4509: gap of 100 bp
* 4510 6325: contig of 1816 bp in length
* 6326 6425: gap of 100 bp
* 6426 8007: contig of 1582 bp in length
* 8008 8107: gap of 100 bp
* 8108 10195: contig of 2088 bp in length
* 10196 10295: gap of 100 bp
* 10296 24900: contig of 14605 bp in length
* 24901 25000: gap of 100 bp
* 25001 46889: contig of 21889 bp in length
* 46890 46989: gap of 100 bp
* 46990 84365: contig of 37376 bp in length
* 84366 84465: gap of 100 bp
* 84466 120489: contig of 36024 bp in length
* 120490 120589: gap of 100 bp
* 120590 171841: contig of 51252 bp in length.

FEATURES

source

1..171841
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-44C14"
/clone_lib="RPC1-11 Human Male BAC"

1..1290

misc_feature

/note="assembly_fragment"

1391..2428

misc_feature

/note="assembly_fragment"

2529..4409

misc_feature

/note="assembly_fragment"

4510..6325

misc_feature

/note="assembly_fragment"

6426..8007

misc_feature

/note="assembly_fragment"

8108..10195

misc_feature

/note="assembly_fragment"

/note="assembly_fragment"

clone_end:SP6

vector_side:left"

10296..24900

misc_feature

/note="assembly_fragment"

25001..46889

misc_feature

/note="assembly_fragment"

46990..84365

misc_feature

/note="assembly_fragment"

84466..120489

misc_feature

/note="assembly_fragment"

clone_end:T7

vector_side:right"

120590..171841

misc_feature

/note="assembly_fragment"

BASE COUNT 48391 a 35633 c 35324 g 51489 t 1004 others
ORIGIN

```
Query Match      0.3%; Score 70; DB 60; Length 171841;
Best Local Similarity 100.0%; Pred. No. 1.4e-26;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10343 ttttgacacagctcgtctgtcaccaggctggagtgcagtggtgcgatcttggctca 10402
|||||
Db 155984 TTTTGACACAGAGTCTGCTGCTGTCCACAGGCTGGAGTGGTGGTGGTCTTGGCTCA 155925

QY 10403 ctcaaacctc 10412
|||||
Db 155924 CTGCAACCTC 155915

RESULT 78
AL162400/c
LOCUS      AL162400      179743 bp      DNA      HTG      20-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-5P4 map p31.3-32.3, ***
SEQUENCING IN PROGRESS ***, 19 unordered pieces.
ACCESSION  AL162400
VERSION     AL162400.5 GI:9796916
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 179743)
AUTHORS   Pavitt,R.
TITLE     Direct Submission
JOURNAL   Submitted (19-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Aug 12, 2000 this sequence version replaced gi:9212944.
COMMENT   ----- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----- Project Information
            Center project name: BA5P4
            ----- Summary Statistics
            Sequencing program: XGAP4; version 4.5
            Chemistry vector: plasmid; L08752; 100% of reads
            Dye-terminator: ET-amersham; 34% of reads Chemistry:
            Dye-terminator Big Dye; 65% of reads
            Consensus quality: 170797 bases at least Q40
            Consensus quality: 173342 bases at least Q30
            Consensus quality: 175125 bases at least Q20
            Insert size: 177943; sum-of-contigs
            Insert size: 173372; 7.5% error; agarose-fp
            Quality coverage: 4.64x in Q20 bases; sum-of-contigs Quality
            coverage: 4.80x in Q20 bases; agarose-fp
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 23993: contig of 23993 bp in length
            * 23994 24093: gap of 100 bp
            * 24094 44611: contig of 20518 bp in length
            * 44612 44711: gap of 100 bp
            * 44712 46762: contig of 2051 bp in length
            * 46763 46862: gap of 100 bp
            * 46863 50027: contig of 3165 bp in length
            * 50028 50127: gap of 100 bp
            * 50128 55974: contig of 5847 bp in length
            * 55975 56074: gap of 100 bp
            * 56075 64453: contig of 8379 bp in length
            * 64454 64554: gap of 100 bp
            * 64554 74354: contig of 9801 bp in length
            *
```

```
* 74355 74454: gap of 100 bp
* 74455 78351: contig of 3897 bp in length
* 78352 78451: gap of 100 bp
* 78452 86399: contig of 7948 bp in length
* 86400 86499: gap of 100 bp
* 86500 92804: contig of 6305 bp in length
* 92805 92904: gap of 100 bp
* 92905 107377: contig of 14473 bp in length
* 107378 107477: gap of 100 bp
* 107478 114160: contig of 6683 bp in length
* 114161 114260: gap of 100 bp
* 114261 120536: contig of 6276 bp in length
* 120537 120636: gap of 100 bp
* 120637 124088: contig of 3452 bp in length
* 124089 124188: gap of 100 bp
* 124189 133546: contig of 9358 bp in length
* 133547 133646: gap of 100 bp
* 133647 170927: contig of 37281 bp in length
* 170928 171027: gap of 100 bp
* 171028 173259: contig of 2232 bp in length
* 173260 173359: gap of 100 bp
* 173360 175833: contig of 2474 bp in length
* 175834 175933: gap of 100 bp
* 175934 179743: contig of 3810 bp in length.
FEATURES
            Location/Qualifiers
            1..179743
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /map="p31.3-32.3"
            /clone="RP11-5P4"
            /clone_lib="RPC1-11.1"
            1..23993
            /note="assembly_fragment:00175"
            fragment_chain:1
            24094..44611
            /note="assembly_fragment:00160"
            fragment_chain:1
            44712..46762
            /note="assembly_fragment:00932"
            fragment_chain:1
            46863..50027
            /note="assembly_fragment:01816"
            fragment_chain:1
            50128..55974
            /note="assembly_fragment:01472"
            fragment_chain:1
            56075..64453
            /note="assembly_fragment:01420"
            fragment_chain:1
            64554..74354
            /note="assembly_fragment:00702"
            fragment_chain:1
            74455..78351
            /note="assembly_fragment:00423"
            fragment_chain:1
            78452..86399
            /note="assembly_fragment:01008"
            fragment_chain:1
            86500..92804
            /note="assembly_fragment:01551"
            fragment_chain:2
            92905..107377
            /note="assembly_fragment:00267"
            fragment_chain:2
            107478..114160
            /note="assembly_fragment:00046"
            114261..120536
            /note="assembly_fragment:00078"
            120637..124088
            /note="assembly_fragment:00944"
            124189..133546
            /note="assembly_fragment:00996"
```

```
misc_feature 133647. .170927
/Note="assembly_fragment:01775"
misc_feature 171028. .173259
/Note="assembly_fragment:01999"
misc_feature 173360. .175833
/Note="assembly_fragment:02016
fragment_chain:3"
misc_feature 175934. .179743
/Note="assembly_fragment:01457
fragment_chain:3
clone_end:T7
vector_side:right"
BASE COUNT 53015 a 36183 c 36336 g 52405 t 1804 others
ORIGIN

Query Match 0.3%; Score 70; DB 79; Length 179743;
Best Local Similarity 100.0%; Pred. No. 1.4e-26;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17556 agctggtctcgaactcctgacccaggtgatccaccacccagctcccaaaagtgtg 17615
|||||
Db 74064 AGCTGCTCGAATCTCGACCTGACCTGAGTCCACCCACCTCAGCTCCCAAAAGTGTG 74005

QY 17616 ggattacagg 17625
|||||
Db 74004 GGATTACAGG 73995

RESULT 79
AC074011
LOCUS AC074011 180462 bp DNA HTG 22-MAR-2001
DEFINITION Homo sapiens chromosome UNK clone RP11-780J6, WORKING DRAFT
SEQUENCE, 1 unordered pieces.
ACCESSION AC074011.4 GI:13431259
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 180462)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Mar 22, 2001 this sequence version replaced gi:11995647.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0780J06
----- Summary Statistics -----
Sequencing vector: M13; 93%
Sequencing vector: plasmid; 7%
Chemistry: Dye-primer ET; 93% of reads
Assembly: Terminator Big Dye; 7% of reads
Assembly program: Phrap; version 0.90319
Consensus quality: 180365 bases at least Q40
Consensus quality: 180457 bases at least Q30
Consensus quality: 180462 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 181641; sum-of-ctnigs
Quality coverage: 7.42 in Q20 bases; agarose-fp
Quality coverage: 7.97 in Q20 bases; sum-of-ctnigs
-----
```

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1. 180462: contig of 180462 bp in length.
FEATURES
source
1. .180462
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-780J6"
misc_feature 1. .180462
/Note="assembly_name:Contig5
clone_end:T7
vector_side:right"
BASE COUNT 48849 a 42940 c 42516 g 46157 t
ORIGIN

Query Match 0.3%; Score 70; DB 75; Length 180462;
Best Local Similarity 100.0%; Pred. No. 1.4e-26;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17555 caggctggtctcgaactcctgacccaggtgatccaccacccagctcccaaaagtgtt 17614
|||||
Db 115740 CAGGCTGGTCTCGAATCTCGACCTGACCTGAGTCCACCCACCTCAGCTCCCAAAAGTGT 115799

QY 17615 gggattacag 17624
|||||
Db 115800 GGGATTACAG 115809

RESULT 80
AC009727
LOCUS AC009727 193159 bp DNA HTG 14-JAN-2001
DEFINITION Homo sapiens chromosome 12 clone RP11-533J15, WORKING DRAFT
SEQUENCE, 12 unordered pieces.
ACCESSION AC009727
VERSION AC009727.4 GI:12083977
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,M., Brown,E., Bryant,N.P., Buhay,C.,
Burck,P., Burckett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Deigado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
```


Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8253

Center clone name: 600_D_20

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 179734 bases at least Q40

Consensus quality: 187831 bases at least Q30

Consensus quality: 190379 bases at least Q20

Insert size: 192000; agarose-fp

Insert size: 191805; sum-of-contigs

Quality coverage: 4.5 in Q20 bases; agarose-fp

Quality coverage: 4.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1387: contig of 1387 bp in length
* 1388 1487: gap of 100 bp
* 1488 2669: contig of 1182 bp in length
* 2670 2769: gap of 100 bp
* 2770 4072: contig of 1303 bp in length
* 4073 4172: gap of 100 bp
* 4173 4518: contig of 346 bp in length
* 4519 4618: gap of 100 bp
* 4619 6850: contig of 2232 bp in length
* 6851 6950: gap of 100 bp
* 6951 8062: contig of 1112 bp in length
* 8063 8162: gap of 100 bp
* 8163 9714: contig of 1552 bp in length
* 9715 9814: gap of 100 bp
* 9815 12197: contig of 2383 bp in length
* 12198 12297: gap of 100 bp
* 12298 14434: contig of 2137 bp in length
* 14435 14534: gap of 100 bp
* 14535 16239: contig of 1705 bp in length
* 16240 16339: gap of 100 bp
* 16340 18999: contig of 2660 bp in length
* 19000 19099: gap of 100 bp
* 19100 20889: contig of 1790 bp in length
* 20890 20989: gap of 100 bp
* 20990 23667: contig of 2678 bp in length
* 23668 23767: gap of 100 bp
* 23768 26382: contig of 2615 bp in length
* 26383 26482: gap of 100 bp
* 26483 31356: contig of 4874 bp in length
* 31357 31456: gap of 100 bp
* 31457 35860: contig of 4404 bp in length
* 35861 35960: gap of 100 bp
* 35961 43693: contig of 7733 bp in length
* 43694 43793: gap of 100 bp
* 43794 52090: contig of 8297 bp in length
* 52091 52190: gap of 100 bp
* 52191 58524: contig of 6334 bp in length
* 58525 58624: gap of 100 bp
* 58625 70938: contig of 12314 bp in length
* 70939 71038: gap of 100 bp
* 71039 85686: contig of 14648 bp in length
* 85687 85786: gap of 100 bp
* 85787 99324: contig of 13538 bp in length
* 99325 99424: gap of 100 bp
* 99425 115316: contig of 15892 bp in length
* 115317 115416: gap of 100 bp

FEATURES

source
1..194405
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone_lib="RPCI-11 Human Male BAC"
/clone="RP11-600D20"
1..1387
/note="assembly_fragment"
1488..2669
/note="assembly_fragment"
2770..4072
/note="assembly_fragment"
4173..4518
/note="assembly_fragment"
clone_end:T7
vector_side:right
4619..6850
/note="assembly_fragment"
6951..8062
/note="assembly_fragment"
8163..9714
/note="assembly_fragment"
9815..12197
/note="assembly_fragment"
12298..14434
/note="assembly_fragment"
14535..16239
/note="assembly_fragment"
16340..18999
/note="assembly_fragment"
19100..20889
/note="assembly_fragment"
20990..23667
/note="assembly_fragment"
23768..26382
/note="assembly_fragment"
26483..31356
/note="assembly_fragment"
31457..35860
/note="assembly_fragment"
35961..43693
/note="assembly_fragment"
43794..52090
/note="assembly_fragment"
52191..58524
/note="assembly_fragment"
clone_end:SP6
vector_side:right
58625..70938
/note="assembly_fragment"
71039..85686
/note="assembly_fragment"
85787..99324
/note="assembly_fragment"
99425..115316
/note="assembly_fragment"
115417..131288
/note="assembly_fragment"
131389..148135
/note="assembly_fragment"
148236..172728
/note="assembly_fragment"
172829..194405
/note="assembly_fragment"

BASE COUNT 56398 a 39349 c 40959 g 55093 t 2606 others
ORIGIN

Query Match 0.3%; Score 70; DB 71; Length 194405;
Best Local Similarity 100.0%; Pred. No. 1.4e-26;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10343 tttagacagagctcgtctcgtccacgagctggagtcagtgagtcgagtcgttcgctca 10402
|||||
Db 177355 TTTTGACACAGAGCTCGCTCTGTACACAGCTGGAGTGGCTGCGATCTGGCTCA 177414
|||||
QY 10403 ctgcaacctc 10412
|||||
Db 177415 CTGCAACCTC 177424

RESULT 82

AC008121

LOCUS

DEFINITION Homo sapiens chromosome 12 clone RP11-407N8, HTG 01-MAR-2001
PROGRESS ***, 95 unordered pieces. *** SEQUENCING IN

AC008121

AC008121.22 GI:13173594

KEYWORDS HTG; HTGS_PHASE1.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 305583)
Muzny,D.M., Adams,C., Amaratunga,H.C., Are,J.R., Banks,T., Barbiana,J.,
Alsbrooks,S.L., Blakenburg,K., Bonin,D., Bouck,J.,
Benton,J., Bieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotot,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisege,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Nowhaton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenad,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 305583)

Worley,K.C.

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (24-JUL-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Mar 1, 2001 this sequence version replaced gi:13162454.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: HMRH

Center clone name: RP11-407N8

----- Summary Statistics

Sequencing vector: Plasmid: M77789

Sequencing vector: M13; L08821

Chemistry: Dye-primer Bodypy: 32% of reads

Chemistry: Dye-terminator Big Dye: 68% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 249729 bases at least Q40

Consensus quality: 283537 bases at least Q30

Consensus quality: 297805 bases at least Q20

Estimated insert size: 262266; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 2.9x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)

* NOTE: This sequence may represent more than one clone.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 95 contigs and their order in this sequence record is

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 14541: contig of 14541 bp in length

* 14542 14641: gap of unknown length

* 14642 25620: contig of 10979 bp in length

* 25621 25720: gap of unknown length

* 25721 37465: contig of 11745 bp in length

* 37466 37565: gap of unknown length

* 37566 48882: contig of 11317 bp in length

* 48883 48982: gap of unknown length

* 48983 60885: contig of 11903 bp in length

* 60886 60985: gap of unknown length

* 60986 69310: contig of 8325 bp in length

* 69311 69410: gap of unknown length

* 69411 78456: contig of 9046 bp in length

* 78457 78556: gap of unknown length

* 78557 87806: contig of 9250 bp in length

* 87807 87906: gap of unknown length

* 87907 96321: contig of 8415 bp in length

* 96322 96421: gap of unknown length

* 96422 102595: contig of 6174 bp in length

* 102596 102695: gap of unknown length

* 102696 109793: contig of 7098 bp in length

* 109794 109893: gap of unknown length

* 109894 117256: contig of 7363 bp in length

* 117257 117356: gap of unknown length

* 117357 122418: contig of 5062 bp in length

* 122419 122518: gap of unknown length

* 122519 128390: contig of 5872 bp in length

* 128391 128490: gap of unknown length

* 128491 136085: contig of 7595 bp in length

* 136086 136185: gap of unknown length

* 136186 140904: contig of 4719 bp in length

* 140905 141004: gap of unknown length

* 141005 146210: contig of 5206 bp in length

* 146211 146310: gap of unknown length

* 146311 150881: contig of 4571 bp in length

* 150882 150981: gap of unknown length

TITLE

JOURNAL

REFERENCE

AUTHORS

* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 705: contig of 705 bp in length
* 805: gap of 100 bp
* 806 1510: contig of 705 bp in length
* 1511 1610: gap of 100 bp
* 1611 2326: contig of 716 bp in length
* 2327 2426: gap of 100 bp
* 2427 3158: contig of 732 bp in length
* 3159 3258: gap of 100 bp
* 3259 3967: contig of 709 bp in length
* 3968 4067: gap of 100 bp
* 4068 4775: contig of 708 bp in length
* 4776 4875: gap of 100 bp
* 4876 5583: contig of 708 bp in length
* 5584 5683: gap of 100 bp
* 5684 6386: contig of 703 bp in length
* 6387 6486: gap of 100 bp
* 6487 7198: contig of 712 bp in length
* 7199 7298: gap of 100 bp
* 7299 7999: contig of 701 bp in length
* 8000 8099: gap of 100 bp
* 8100 8830: contig of 731 bp in length
* 8831 8930: gap of 100 bp
* 8931 9655: contig of 725 bp in length
* 9656 9755: gap of 100 bp
* 9756 10465: contig of 710 bp in length
* 10466 10565: gap of 100 bp
* 10566 11274: contig of 709 bp in length
* 11275 11374: gap of 100 bp
* 11375 12087: contig of 713 bp in length
* 12088 12187: gap of 100 bp
* 12188 12913: contig of 726 bp in length
* 12914 13013: gap of 100 bp
* 13014 13737: contig of 724 bp in length
* 13738 13837: gap of 100 bp
* 13838 14544: contig of 707 bp in length
* 14545 14644: gap of 100 bp
* 14645 15369: contig of 725 bp in length
* 15370 15469: gap of 100 bp
* 15470 16198: contig of 729 bp in length
* 16199 16298: gap of 100 bp
* 16299 17080: contig of 782 bp in length
* 17081 17180: gap of 100 bp
* 17181 17889: contig of 709 bp in length
* 17890 17989: gap of 100 bp
* 17990 18699: contig of 710 bp in length
* 18700 18799: gap of 100 bp
* 18800 19513: contig of 714 bp in length
* 19514 19613: gap of 100 bp
* 19614 20336: contig of 723 bp in length
* 20337 20436: gap of 100 bp
* 20437 21162: contig of 726 bp in length
* 21163 21262: gap of 100 bp
* 21263 21958: contig of 696 bp in length
* 21959 22058: gap of 100 bp
* 22059 22772: contig of 714 bp in length
* 22773 22872: gap of 100 bp
* 22873 23576: contig of 704 bp in length
* 23577 23676: gap of 100 bp
* 23677 24404: contig of 728 bp in length
* 24405 24504: gap of 100 bp
* 24505 25231: contig of 727 bp in length
* 25232 25331: gap of 100 bp
* 25332 26052: contig of 721 bp in length
* 26053 26152: gap of 100 bp
* 26153 26869: contig of 717 bp in length
* 26870 26969: gap of 100 bp
* 26970 27670: contig of 701 bp in length
* 27671 27770: gap of 100 bp

* 27771 28466: contig of 696 bp in length
* 28467 28566: gap of 100 bp
* 28567 29282: contig of 716 bp in length
* 29283 29382: gap of 100 bp
* 29383 30109: contig of 727 bp in length
* 30110 30209: gap of 100 bp
* 30210 30920: contig of 711 bp in length
* 30921 31020: gap of 100 bp
* 31021 31734: contig of 714 bp in length
* 31735 31834: gap of 100 bp
* 31835 32555: contig of 721 bp in length
* 32556 32655: gap of 100 bp
* 32656 33369: contig of 714 bp in length
* 33370 33469: gap of 100 bp
* 33470 34167: contig of 698 bp in length
* 34168 34267: gap of 100 bp
* 34268 34979: contig of 712 bp in length
* 34980 35079: gap of 100 bp
* 35080 35786: contig of 707 bp in length
* 35787 35886: gap of 100 bp
* 35887 36587: contig of 701 bp in length
* 36588 36687: gap of 100 bp
* 36688 37417: contig of 730 bp in length
* 37418 37517: gap of 100 bp
* 37518 38223: contig of 706 bp in length
* 38224 38323: gap of 100 bp
* 38324 39008: contig of 685 bp in length
* 39009 39108: gap of 100 bp
* 39109 39819: contig of 711 bp in length
* 39820 39919: gap of 100 bp
* 39920 40631: contig of 712 bp in length
* 40632 40731: gap of 100 bp
* 40732 41439: contig of 708 bp in length
* 41440 41539: gap of 100 bp
* 41540 42261: contig of 722 bp in length
* 42262 42361: gap of 100 bp
* 42362 43084: contig of 723 bp in length
* 43085 43184: gap of 100 bp
* 43185 43907: contig of 723 bp in length
* 43908 44007: gap of 100 bp
* 44008 44723: contig of 716 bp in length
* 44724 44823: gap of 100 bp
* 44824 45543: contig of 720 bp in length
* 45544 45643: gap of 100 bp
* 45644 46372: contig of 729 bp in length
* 46373 46472: gap of 100 bp
* 46473 47187: contig of 715 bp in length
* 47188 47287: gap of 100 bp
* 47288 47998: contig of 711 bp in length
* 47999 48098: gap of 100 bp
* 48099 48799: contig of 701 bp in length
* 48800 48999: gap of 100 bp
* 48900 49618: contig of 719 bp in length
* 49619 49718: gap of 100 bp
* 49719 50433: contig of 715 bp in length
* 50434 50533: gap of 100 bp
* 50534 51252: contig of 719 bp in length
* 51253 51352: gap of 100 bp
* 51353 52078: contig of 726 bp in length
* 52079 52178: gap of 100 bp
* 52179 52889: contig of 711 bp in length
* 52890 52989: gap of 100 bp
* 52990 53719: contig of 730 bp in length
* 53720 53819: gap of 100 bp
* 53820 54551: contig of 732 bp in length
* 54552 54651: gap of 100 bp
* 54652 55382: contig of 731 bp in length
* 55383 55482: gap of 100 bp
* 55483 56183: contig of 701 bp in length
* 56184 56283: gap of 100 bp

Query Match
Best Local Similarity

0.3%; Score 69; DB 78; Length 60280;
100.0%; Pred. No. 4.4e-26;

Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12635 aattttatttttagtagacaggggtttctcacgttgctcaggtgctcacaactc 12694
 |||||
 Db 473 ANTATTGATTATTAGTAGACAGCGGGTTCTCCACGTTGTCAGGCTGGTCTCAAACTC 532
 |||||

QY 12695 ctgacctca 12703
 |||||
 Db 533 CTGACCTCA 541
 |||||

RESULT 84
 HUAC002551
 LOCUS
 DEFINITION Human Chromosome 16 BAC clone CIT987SK-A-951C11, complete sequence.
 ACCESSION AC002551
 VERSION AC002551.1 GI:2809276
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 114411)
 Loftus,B.J., Kim,U.J., Sneddon,V.P., Kalush,F., Brandon,R.,
 Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,
 Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S.,
 Eichler,E.E., Harris,P.C., Venter,J.C. and Adams,M.D.
 Genome duplications and other features in 12 Mb of DNA sequence
 from human chromosome 16p and 16q
 Genomics 60 (3), 295-308 (1999)
 99425270
 10493829
 2 (bases 1 to 114411)
 Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
 Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
 Human Chromosome 16 BAC clone CIT987SK-A-951C11
 Unpublished
 3 (bases 1 to 114411)
 Adams,M.D.
 Direct Submission
 Submitted (19-SEP-1997) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 4 (bases 1 to 114411)
 Adams,M.D.
 Direct Submission
 Submitted (26-JAN-1998) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 5 (bases 1 to 114411)
 Adams,M.D.
 Direct Submission
 Submitted (28-JAN-1998) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 6 (bases 1 to 114411)
 Adams,M.D.
 Direct Submission
 Submitted (05-FEB-1998) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 7 (bases 1 to 114411)
 Adams,M.D. and Loftus,B.J.
 Direct Submission
 Submitted (17-MAR-1998) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 On Jan 26, 1998 this sequence version replaced gi:2642422.
 Address all correspondence to: Mark Adams The Institute for Genomic
 Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
 address: mdadams@tigr.org. The orientation of the sequence is
 from SP6 end to T7 end. Genes were identified by a combination of
 five methods including: XGAP (available by anonymous ftp from
 arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
 Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/>)
 chris@GENSCANW.html) searches of the complete sequence against a
 peptide database, and the human gene Index database at TIGR
 (<http://www.tigr.org/tdb/hg1/hgi.html>). Genes without pepetide

homology having spliced EST hits are termed 'Unknown gene product'.
 Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>).

FEATURES
 source
 1. 114411
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16p12.1"
 /clone="A-951C11"
 complement(<3210..3493)
 /gene="A-951C1A1."
 complement(<3210..3493)
 /gene="A-951C1A1."
 complement(<3210..3430)
 /gene="A-951C1A1."
 /codon_start=1
 /product="Transcription factor (TFIIIC) alpha chain, (3'
 partial)"
 /protein_id="AAC05811.1"
 /db_xref="GI:2833636"
 /translation="MDALESLLDEVALEGLDGLCLPALWSRLRTRVPPPLPLEPCTQ
 EFLWRLATHPGISFYEEPRERPDQLQDR"
 6327..6444
 /db_xref="dbSTS:G02122"
 101138..101255
 /db_xref="dbSTS:G02122"
 29730 a 26699 c 27271 g 30711 t
 BASE COUNT
 ORIGIN

Query Match 0.3%; Score 69; DB 97; Length 114411;
 Best Local Similarity 100.0%; Pred. No. 4.8e-26;
 Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17383 ttttcccgagctggagtgcaatggcggtgctcagctcagctcagctccgctccg 17442
 |||||
 Db 28140 TGTGTCGAGGCTGAGTGCAATGGGTGATCTCAGCTCAGCTGCAACCTCCGCTCCCG 28199
 |||||

QY 17443 gtccaagca 17451
 |||||
 Db 28200 GTTCAAGCA 28208
 |||||

RESULT 85
 AC011979/c
 LOCUS
 DEFINITION Homo sapiens clone RP11-16F15, WORKING DRAFT SEQUENCE, 30 unordered
 pieces.
 ACCESSION AC011979
 VERSION AC011979.3 GI:7341927
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 137641)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens, clone RP11-16F15
 Unpublished
 2 (bases 1 to 137641)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhalter,B.,
 Brown,A., Castle,A., Collinge,M., Collins,S., Collymore,A.,
 Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Kerton,J.,
 Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meidrim,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,


```
misc_feature 58426. .64502
/note="assembly_fragment"
misc_feature 64603. .72223
/note="assembly_fragment"
misc_feature 72324. .79639
/note="assembly_fragment"
misc_feature 79740. .88137
/note="assembly_fragment"
misc_feature 88238. .96553
/note="assembly_fragment"
misc_feature 96654. .105243
/note="assembly_fragment"
misc_feature 105344. .113380
/note="assembly_fragment"
misc_feature 113481. .122965
/note="assembly_fragment
clone_end:T7
vector_side:right"
misc_feature 123066. .137641
/note="assembly_fragment"

BASE COUNT 39168 a 29416 c 27916 g 38239 t 2902 others
ORIGIN

Query Match 0.3%; Score 69; DB 62; Length 137641;
Best Local Similarity 100.0%; Pred. No. 4.9e-26;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17379 ctctgttgcaggctgagtgcaatgcgtatctcagctcactgcacactcgctc 17438
|||||
Db 135077 CTCTGTGTGCCAGCTGAGTGCATGCGTGCATCTCAGCTCACTGCAACCTCGCGCTC 135018

Qy 17439 ccgggttca 17447
|||||
Db 135017 CCGGTTCA 135009

RESULT 86
AC026709/c 153620 bp DNA HTG 31-AUG-2000
LOCUS Homo sapiens chromosome 5 clone CTD-2084I118, WORKING DRAFT
DEFINITION
ACCESSION AC026709
VERSION AC026709.4 GI:9954678
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 153620)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 31, 2000 this sequence version replaced gi:8655954.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 660015
Center clone name: CITB-HL_2084I18
-----
Summary Statistics
Consensus quality: 146148 bases at least Q40
Consensus quality: 150906 bases at least Q30
Consensus quality: 151921 bases at least Q20
Estimated insert size: 147000; pulse field gel estimation

Estimated insert size: 153120; sum-of-contigs estimation
Quality coverage: 3.94 in Q20 bases; pulse field gel estimation
Quality coverage: 3.78 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1
3214 3213: contig of 3213 bp in length
3314 8336: gap of unknown length
3314 8336: contig of 5023 bp in length
8337 8436: gap of unknown length
8437 17282: contig of 8846 bp in length
17283 17382: gap of unknown length
17383 35950: contig of 18568 bp in length
35951 36050: gap of unknown length
36051 48131: contig of 12081 bp in length
48132 48231: gap of unknown length
48232 97104: contig of 48873 bp in length
97105 97204: gap of unknown length
97205 104494: contig of 7290 bp in length
104495 104594: gap of unknown length
104595 109570: contig of 4976 bp in length
109571 109670: gap of unknown length
109671 138178: contig of 28508 bp in length
138179 138278: gap of unknown length
138279 139698: contig of 1420 bp in length
139699 139798: gap of unknown length
139799 143196: contig of 3398 bp in length
143197 143296: gap of unknown length
143297 153620: contig of 10324 bp in length.

FEATURES
Location/Qualifiers
1. 153620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone_lib="CalTech human BAC library D"
BASE COUNT 42405 a 35733 c 35061 g 39321 t 1100 others
ORIGIN

Query Match 0.3%; Score 69; DB 70; Length 153620;
Best Local Similarity 100.0%; Pred. No. 5e-26;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8915 atctggctcactgcagctccacctccaggttcaagtgcattctctcctcagctcc 8974
|||||
Db 100277 ATCTCGGCTCACTGCAGCTCCACCTCCAGGTCAAGTGATTCCTCCTCAGCTCC 100218

Qy 8975 caagtacgt 8983
|||||
Db 100217 CAAGTAGCT 100209

RESULT 87
AL357046/c 154577 bp DNA HTG 23-JAN-2001
LOCUS Homo sapiens chromosome 6 clone RP11-689K9, *** SEQUENCING IN
DEFINITION PROGRESS ***, 24 unordered pieces.
ACCESSION AL357046
VERSION AL357046.3 GI:9863806
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 154577)
AUTHORS Burton, J.
```

TITLE
JOURNAL

Direct Submission
Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:9214048.

COMMENT

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA689K9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 138465 bases at least Q40
Consensus quality: 144377 bases at least Q30
Consensus quality: 148251 bases at least Q20
Insert size: 152277; sum-of-contigs
Insert size: 183673; agarose-fp
Quality coverage: 3.10x in Q20 bases; sum-of-contigs Quality
coverage: 2.91x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 20042: contig of 20042 bp in length
* 20043 20142: gap of 100 bp
* 20143 28431: contig of 8289 bp in length
* 28432 28531: gap of 100 bp
* 28532 31271: contig of 2740 bp in length
* 31272 31371: gap of 100 bp
* 31372 36284: contig of 4913 bp in length
* 36285 36384: gap of 100 bp
* 36385 43000: contig of 6616 bp in length
* 43001 43100: gap of 100 bp
* 43101 53686: contig of 10586 bp in length
* 53687 53786: gap of 100 bp
* 53787 57104: contig of 3318 bp in length
* 57105 57204: gap of 100 bp
* 57205 60355: contig of 3151 bp in length
* 60356 60455: gap of 100 bp
* 60456 84981: contig of 24526 bp in length
* 84982 85081: gap of 100 bp
* 85082 91789: contig of 6708 bp in length
* 91790 91889: gap of 100 bp
* 91890 96630: contig of 4741 bp in length
* 96631 96730: gap of 100 bp
* 96731 103157: contig of 6427 bp in length
* 103158 103257: gap of 100 bp
* 103258 106192: contig of 2935 bp in length
* 106193 106292: gap of 100 bp
* 106293 108453: contig of 2161 bp in length
* 108454 108553: gap of 100 bp
* 108554 112185: contig of 3632 bp in length
* 112186 112285: gap of 100 bp
* 112286 117548: contig of 5263 bp in length
* 117549 117648: gap of 100 bp
* 117649 120443: contig of 2495 bp in length
* 120144 120243: gap of 100 bp
* 120244 127374: contig of 7131 bp in length
* 127375 127474: gap of 100 bp
* 127475 131033: contig of 3559 bp in length
* 131034 131133: gap of 100 bp
* 131134 134291: contig of 3158 bp in length
* 134292 134391: gap of 100 bp
* 134392 137558: contig of 3167 bp in length

* 137559 137658: gap of 100 bp
* 137659 144322: contig of 6664 bp in length
* 144323 144422: gap of 100 bp
* 144423 147469: contig of 3047 bp in length
* 147470 147569: gap of 100 bp
* 147570 154577: contig of 7008 bp in length.
FEATURES
Location/Qualifiers
1. .154577
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="Rp11-689K9"
/clone_lib="RPCI-11.3"
1. .20042
/note="assembly_fragment:00346
fragment_chain:1"
misc_feature
20143. .28431
/note="assembly_fragment:01178
fragment_chain:1"
misc_feature
28532. .31271
/note="assembly_fragment:00155
fragment_chain:1"
misc_feature
31372. .36284
/note="assembly_fragment:00003
fragment_chain:1"
misc_feature
36385. .43000
/note="assembly_fragment:01172
fragment_chain:1"
misc_feature
43101. .53686
/note="assembly_fragment:00257
fragment_chain:2"
misc_feature
53787. .57104
/note="assembly_fragment:00918
fragment_chain:2"
misc_feature
57205. .60355
/note="assembly_fragment:00351
fragment_chain:2"
misc_feature
60456. .84981
/note="assembly_fragment:00434
fragment_chain:3"
misc_feature
85082. .91789
/note="assembly_fragment:00028
fragment_chain:3"
misc_feature
91890. .96630
/note="assembly_fragment:00419
fragment_chain:3"
misc_feature
96731. .103157
/note="assembly_fragment:01056
fragment_chain:4"
misc_feature
103258. .106192
/note="assembly_fragment:00093
fragment_chain:4"
misc_feature
106293. .108453
/note="assembly_fragment:00987
fragment_chain:4"
misc_feature
108554. .112185
/note="assembly_fragment:00232
fragment_chain:5"
misc_feature
112286. .117548
/note="assembly_fragment:00328
fragment_chain:5"
misc_feature
117649. .120143
/note="assembly_fragment:00179"
fragment_chain:5"
misc_feature
120244. .127374
/note="assembly_fragment:00180"
fragment_chain:5"
misc_feature
127475. .131033
/note="assembly_fragment:00309"
fragment_chain:5"
misc_feature
131134. .134291
/note="assembly_fragment:00432"
fragment_chain:5"
misc_feature
134392. .137558
/note="assembly_fragment:00660"
fragment_chain:5"
misc_feature
137659. .144322
/note="assembly_fragment:00716"
fragment_chain:5"

* 125129 137367: contig of 12239 bp in length
* 137368 137467: gap of 100 bp
* 137468 153264: contig of 15797 bp in length
* 153265 153364: gap of 100 bp
* 153365 168111: contig of 14747 bp in length.

FEATURES

```
source
1. .168111
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-105K20"
/clone_lib="RPC1-11 Human Male BAC"
1. .426
/note="assembly_fragment"
clone_end:SP6
vector_side:left
527. .1891
/note="assembly_fragment"
1992. .2994
/note="assembly_fragment"
3095. .4450
/note="assembly_fragment"
4551. .5934
/note="assembly_fragment"
6035. .8109
/note="assembly_fragment"
8210. .10120
/note="assembly_fragment"
10221. .13572
/note="assembly_fragment"
13673. .17728
/note="assembly_fragment"
17829. .21346
/note="assembly_fragment"
21447. .25332
/note="assembly_fragment"
25433. .29242
/note="assembly_fragment"
29343. .32324
/note="assembly_fragment"
32425. .36856
/note="assembly_fragment"
36957. .41393
/note="assembly_fragment"
41494. .46862
/note="assembly_fragment"
46963. .52806
/note="assembly_fragment"
52907. .58548
/note="assembly_fragment"
58649. .64705
/note="assembly_fragment"
64806. .70785
/note="assembly_fragment"
70886. .78894
/note="assembly_fragment"
78995. .87124
/note="assembly_fragment"
87225. .95140
/note="assembly_fragment"
95241. .103970
/note="assembly_fragment"
104071. .111902
/note="assembly_fragment"
112003. .125028
/note="assembly_fragment"
125129. .137367
/note="assembly_fragment"
137468. .153264
/note="assembly_fragment"
153365. .168111
/note="assembly_fragment"
```

```
clone_end:T7
vector_side:right"
BASE COUNT 49155 a 34309 c 33626 g 48205 t 2816 others
ORIGIN
Query Match 0.3%; Score 69; DB 69; Length 168111;
Best Local Similarity 100.0%; Pred. No. 5; ie-26;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 17379 ctctgtgtgccagctggagtcgaatggcgtgatcctcagctcgaacctcgccctc 17438
|||||
Db 68538 CTCCTGTGCCCCAGCTGGAGTGCATGGCGTGATCTCAGCTCACTGCACACCTCGCCCTC 68479
|||||
Qy 17439 ccgggttca 17447
|||||
Db 68478 CCGGTTCA 68470

RESULT 89
AC023388/c
LOCUS AC023388 168584 bp DNA HTG 01-MAR-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-35B18 map 18, WORKING DRAFT
SEQUENCE, 19 unordered pieces.
AC023388
AC023388.2 GI:7139714
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 168584)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP11-35B18
Unpublished
2 (bases 1 to 168584)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bada,F., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campolano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,T.M.,
Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 1, 2000 this sequence version replaced gi:6970538.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5335
Center clone name: 35_B18
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
```

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 157982 bases at least Q40
Consensus quality: 163050 bases at least Q30
Consensus quality: 165225 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 166784; sum-of-contigs
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1743: contig of 1743 bp in length
* 1744 1843: gap of 100 bp
* 1844 3703: contig of 1860 bp in length
* 3704 3803: gap of 100 bp
* 3804 6108: contig of 2305 bp in length
* 6109 6208: gap of 100 bp
* 6209 10925: contig of 4717 bp in length
* 10926 11025: gap of 100 bp
* 11026 13874: contig of 2849 bp in length
* 13875 13974: gap of 100 bp
* 13975 18361: contig of 4387 bp in length
* 18362 18461: gap of 100 bp
* 18462 23451: contig of 4990 bp in length
* 23452 23551: gap of 100 bp
* 23552 28686: contig of 5135 bp in length
* 28687 28786: gap of 100 bp
* 28787 34325: contig of 5539 bp in length
* 34326 34425: gap of 100 bp
* 34426 40238: contig of 5813 bp in length
* 40239 40338: gap of 100 bp
* 40339 46534: contig of 6196 bp in length
* 46535 46634: gap of 100 bp
* 46635 56056: contig of 9422 bp in length
* 56057 56156: gap of 100 bp
* 56157 69854: contig of 13698 bp in length
* 69855 69954: gap of 100 bp
* 69955 82190: contig of 12236 bp in length
* 82191 82290: gap of 100 bp
* 82291 97895: contig of 15605 bp in length
* 97896 97995: gap of 100 bp
* 97996 115061: contig of 17066 bp in length
* 115062 115161: gap of 100 bp
* 115162 130139: contig of 14978 bp in length
* 130140 130239: gap of 100 bp
* 130240 147825: contig of 17586 bp in length
* 147826 147925: gap of 100 bp
* 147926 168584: contig of 20659 bp in length.

FEATURES

Source	1..168584
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="18"
	/map="18"
	/clone_lib="RPCI-11 Human Male BAC"
	1..1743
misc_feature	/note="assembly_fragment"
misc_feature	1844..3703
misc_feature	/note="assembly_fragment"
misc_feature	3804..6108
misc_feature	/note="assembly_fragment"
misc_feature	6209..10925
misc_feature	/note="assembly_fragment"
misc_feature	11026..13874
misc_feature	/note="assembly_fragment"

FEATURES	Source
LOCUS	AC021178 171901 bp DNA HTG 07-JUL-2000
DEFINITION	Homo sapiens chromosome 5 clone RP11-699022, WORKING DRAFT
KEYWORDS	SEQUENCE, 24 unordered pieces.
VERSION	AC021178
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 171901)
AUTHORS	Waterston,R.H.
TITLE	The sequence of Homo sapiens clone
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 171901)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (14-JAN-2000) Genome Sequencing Center, Washington
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,
	MO 63108, USA
COMMENT	On Jun 17, 2000 this sequence version replaced gi:7272322.

misc_feature	13975..18361
	/note="assembly_fragment"
	clone_end:77
	vector_side:left
misc_feature	18462..23451
	/note="assembly_fragment"
misc_feature	23552..28686
	/note="assembly_fragment"
	clone_end:SP6
	vector_side:left
misc_feature	28787..34325
	/note="assembly_fragment"
misc_feature	34426..40238
	/note="assembly_fragment"
misc_feature	40339..46534
	/note="assembly_fragment"
misc_feature	46635..56056
	/note="assembly_fragment"
misc_feature	56157..69854
	/note="assembly_fragment"
misc_feature	69955..82190
	/note="assembly_fragment"
misc_feature	82291..97895
	/note="assembly_fragment"
misc_feature	97996..115061
	/note="assembly_fragment"
misc_feature	115162..130139
	/note="assembly_fragment"
misc_feature	130240..147825
	/note="assembly_fragment"
misc_feature	147926..168584
	/note="assembly_fragment"
BASE COUNT	43714 a 38724 c 38852 g 45489 t 1805 others
ORIGIN	

Query Match	0.3%; Score 69; DB 68; Length 168584;
Best Local Similarity	100.0%; Pred. No. 5,1e-26;
Matches	69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	8915	atctggctactgcagctccacccctccaggttcacgttcctcctcgcctcagctcc	8974
DB	43223	ATCTCGGCTCACTGCAGCTCCACCTCCAGGTTCAAGTATTCCTCCTCAGCTCC	43164
QY	8975	caagtagct	8983
DB	43163	CAAGTAGCT	43155

RESULT	90
LOCUS	AC021178/c
DEFINITION	Homo sapiens chromosome 5 clone RP11-699022, WORKING DRAFT
KEYWORDS	SEQUENCE, 24 unordered pieces.
VERSION	AC021178
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 171901)
AUTHORS	Waterston,R.H.
TITLE	The sequence of Homo sapiens clone
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 171901)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (14-JAN-2000) Genome Sequencing Center, Washington
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,
	MO 63108, USA
COMMENT	On Jun 17, 2000 this sequence version replaced gi:7272322.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml

----- Project Information -----
Center project name: H_NH0699022
----- Summary Statistics -----

Sequencing vector: plasmid; 17%
Chemistry: Dye-primer ET; 83% of reads
Chemistry: Dye-terminator Big Dye; 17% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 156170 bases at least Q40
Consensus quality: 161910 bases at least Q30
Consensus quality: 164825 bases at least Q20
Insert size: 179000; agarose-fp
Quality coverage: 169601; sum-of-contigs
Quality coverage: 3.44 in Q20 bases; agarose-fp
Quality coverage: 3.68 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2277: contig of 2277 bp in length
* 2278: contig of unknown length
* 2378: contig of 3002 bp in length
* 5380: 5479: gap of unknown length
* 5480: 7909: contig of 2430 bp in length
* 7910: 8099: gap of unknown length
* 8010: 10174: contig of 2165 bp in length
* 10175: 10274: gap of unknown length
* 10275: 12835: contig of 2561 bp in length
* 12836: 12935: gap of unknown length
* 12936: 16899: contig of 3964 bp in length
* 16900: 16999: gap of unknown length
* 17000: 21455: contig of 4456 bp in length
* 21456: 21555: gap of unknown length
* 21556: 26110: contig of 4555 bp in length
* 26111: 26210: gap of unknown length
* 26211: 33068: contig of 6858 bp in length
* 33069: 33169: gap of unknown length
* 33169: 38569: contig of 5400 bp in length
* 38569: 38669: gap of unknown length
* 38669: 44587: contig of 5919 bp in length
* 44588: 44688: gap of unknown length
* 44688: 51751: contig of 7064 bp in length
* 51752: 51851: gap of unknown length
* 51852: 58654: contig of 6803 bp in length
* 58655: 58755: gap of unknown length
* 58755: 67294: contig of 8540 bp in length
* 67295: 73394: gap of unknown length
* 73395: 73872: contig of 6478 bp in length
* 73873: 73973: gap of unknown length
* 73973: 81103: contig of 7131 bp in length
* 81104: 88841: contig of 7638 bp in length
* 88842: 88941: gap of unknown length
* 88942: 96513: contig of 7572 bp in length
* 96514: 96613: gap of unknown length
* 96614: 105579: contig of 8965 bp in length
* 105579: 105679: gap of unknown length
* 105679: 115769: contig of 10090 bp in length
* 115769: 115869: gap of unknown length
* 115869: 127456: contig of 11588 bp in length
* 127457: 127557: gap of unknown length
* 127557: 139387: contig of 11830 bp in length
* 139387: 139487: gap of unknown length
* 139487: 153369: contig of 13883 bp in length

* 153370 153469: gap of unknown length
* 153470 171901: contig of 18432 bp in length.

FEATURES

Location/Qualifiers

source

1..171901
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-699022"
1..2277
/note="assembly_name:Contig10"
2378..5379
/note="assembly_name:Contig11"
5480..7909
/note="assembly_name:Contig12"
8010..10174
/note="assembly_name:Contig13"
10275..12835
/note="assembly_name:Contig14"
12936..16899
/note="assembly_name:Contig15"
17000..21455
/note="assembly_name:Contig16"
21556..26110
/note="assembly_name:Contig17"
26211..33068
/note="assembly_name:Contig18"
33169..38568
/note="assembly_name:Contig19"
38669..44587
/note="assembly_name:Contig20"
44688..51751
/note="assembly_name:Contig21"
51852..58654
/note="assembly_name:Contig22"
58755..67294
/note="assembly_name:Contig23"
67395..73872
/note="assembly_name:Contig24"
73973..81103
/note="assembly_name:Contig25"
81204..88841
/note="assembly_name:Contig26"
88942..96513
/note="assembly_name:Contig27"
96614..105578
/note="assembly_name:Contig28"
105679..115768
/note="assembly_name:Contig29"
115869..127456
/note="assembly_name:Contig30"
127557..139386
/note="assembly_name:Contig31"
139487..153369
/note="assembly_name:Contig32"
clone_end:SP6
vector_side:left
153470..171901
/note="assembly_name:Contig33"

BASE COUNT 45428 a 38998 c 39575 g 45592 t 2308 others
ORIGIN

Query Match 0.3%; Score 69; DB 66; Length 171901;
Best Local Similarity 100.0%; Pred.No. 5.1e-26; Indels 0; Gaps 0;

Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8915 attctcggtcactgcagctccaccctccaggttcgaagtattctctgctcagctcc 8974
|||||

Db 13718 ATCTGGCTACCTGCAGCTCCACCTCCAGGTTCAAGTATTCTCTGCTCAGCTCC 13659

QY 8975 caagttagct 8983
|||||

Db 13658 CAAGTAGCT 13650

RESULT 91.

AC037481 177129 bp DNA HTG 12-MAY-2000
LOCUS Homo sapiens chromosome 18 clone RP11-586H24 map 18, WORKING DRAFT
DEFINITION SEQUENCE, 25 unordered pieces.
AC037481
ACCESSION AC037481.2 GI:7770589
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 177129)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 18, clone RP11-586H24
PUBLISHED
2 (bases 1 to 177129)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczeky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 12, 2000 this sequence version replaced gi:7528389.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9297
Center clone name: 586.H.24
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161642 bases at least Q40
Consensus quality: 169887 bases at least Q30
Consensus quality: 172969 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 174729; sum-of-ctgts
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.6 in Q20 bases; sum-of-ctgts

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.
* 1 1124: contig of 1124 bp in length
* 1125 1224: gap of 100 bp
* 1225 2471: contig of 1247 bp in length
* 2472 2571: gap of 100 bp
* 2572 4603: contig of 2032 bp in length
* 4604 4703: gap of 100 bp
* 4704 6664: contig of 1961 bp in length
* 6665 6764: gap of 100 bp
* 6765 8332: contig of 1568 bp in length
* 8333 8432: gap of 100 bp
* 8433 11750: contig of 3318 bp in length
* 11751 11850: gap of 100 bp
* 11851 14880: contig of 3030 bp in length
* 14881 14980: gap of 100 bp
* 14981 17659: contig of 2679 bp in length
* 17660 17759: gap of 100 bp
* 17760 21978: contig of 4219 bp in length
* 21979 22078: gap of 100 bp
* 22079 25080: contig of 3002 bp in length
* 25081 25180: gap of 100 bp
* 25181 28260: contig of 3080 bp in length
* 28261 28360: gap of 100 bp
* 28361 32906: contig of 4546 bp in length
* 32907 33006: gap of 100 bp
* 33007 36950: contig of 3944 bp in length
* 36951 37050: gap of 100 bp
* 37051 41510: contig of 4460 bp in length
* 41511 41610: gap of 100 bp
* 41611 47575: contig of 5965 bp in length
* 47576 47675: gap of 100 bp
* 47676 53350: contig of 5675 bp in length
* 53351 53450: gap of 100 bp
* 53451 58229: contig of 5279 bp in length
* 58230 58229: gap of 100 bp
* 58230 69023: contig of 10194 bp in length
* 69024 69123: gap of 100 bp
* 69124 77125: contig of 8002 bp in length
* 77126 77225: gap of 100 bp
* 77226 86639: contig of 9414 bp in length
* 86640 86739: gap of 100 bp
* 86740 98330: contig of 11591 bp in length
* 98331 98430: gap of 100 bp
* 98431 113678: contig of 15248 bp in length
* 113679 113778: gap of 100 bp
* 113779 126163: contig of 12385 bp in length
* 126164 126263: gap of 100 bp
* 126264 150533: contig of 24270 bp in length
* 150534 150633: gap of 100 bp
* 150634 177129: contig of 26496 bp in length.
FEATURES
Location/Qualifiers
1..177129
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-586H24"
/clone_lib="RPC1-11 Human Male BAC"
1..1124
/note="assembly_fragment"
1225..2471
/note="assembly_fragment"
2572..4603
/note="assembly_fragment"
4704..6664
/note="assembly_fragment"
6765..8332
/note="assembly_fragment"
8433..11750
/note="assembly_fragment"
11851..14880
/note="assembly_fragment"

```

misc_feature 14981..17659
/note="assembly_fragment"
misc_feature 17760..21978
/note="assembly_fragment"
misc_feature 22079..25080
/note="assembly_fragment"
misc_feature 25181..28260
/note="assembly_fragment"
misc_feature 28361..32906
/note="assembly_fragment"
misc_feature 33007..36950
/note="assembly_fragment"
misc_feature 37051..41510
/note="assembly_fragment"
misc_feature 41611..47575
/note="assembly_fragment"
misc_feature 47676..53350
/note="assembly_fragment"
misc_feature 53451..58729
/note="assembly_fragment"
clone_end:SP6
vector_side:left
58830..69023
/note="assembly_fragment"
69124..77125
/note="assembly_fragment"
77226..86639
/note="assembly_fragment"
86740..98330
/note="assembly_fragment"
98431..113678
/note="assembly_fragment"
113779..126163
/note="assembly_fragment"
126264..150533
/note="assembly_fragment"
clone_end:T7
vector_side:left
150634..177129
/note="assembly_fragment"

BASE COUNT 58006 a 30792 c 30462 g 55467 t 2402 others
ORIGIN

Query Match 0.3% Score 69 DB 71 Length 177129;
Best Local Similarity 100.0% Pred.No. 5.1e-26;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17382 ttgtgccaggctggagtgcaatggcgtgatctcagctcaactgcaacctccgcctccg 17441
|||||
Db 92697 TTGTTGCCAGGCTGGAGTGCAATGGCGTGATCTCAGCTCAGTCACTCGACCTCCGCTCCG 92756

QY 17442 ggttcaagc 17450
|||||
Db 92757 GGTTCAGC 92765

RESULT 92
AC009786 179947 bp DNA HTG 23-APR-2000
LOCUS Homo sapiens clone RP11-44J9, WORKING DRAFT SEQUENCE, 4 unordered
DEFINITION pieces.
ACCESSION AC009786
VERSION AC009786.2 GI:7637769
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179947)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-44J9
JOURNAL Unpublished

```

```

REFERENCE
AUTHORS
2 (bases 1 to 179947)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Depayre,E., Devou,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (01-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 23, 2000 this sequence version replaced gi:5815565.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1076
Center clone name: 44.J.9
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-primer-amersham; 3% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 177174 bases at least Q40
Consensus quality: 178432 bases at least Q30
Consensus quality: 178928 bases at least Q20
Insert size: 180000; agarose-fp
Quality coverage: 179647; sum-of-contigs
Quality coverage: 7.3 in Q20 bases; agarose-fp
Quality coverage: 7.3 in Q20 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1529: contig of 1529 bp in length
* 1530 1629: gap of 100 bp
* 1630 27837: contig of 26208 bp in length
* 27838 27937: gap of 100 bp
* 27938 81075: contig of 53138 bp in length
* 81076 81175: gap of 100 bp
* 81176 179947: contig of 98772 bp in length.
FEATURES
source
Location/Qualifiers
1..179947
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-44J9"
/clone_lib="RPCI-11 Human Male BAC"
1..1529
/note="assembly_fragment"
clone_end:T7
vector_side:right
1630..27837
/note="assembly_fragment"
clone_end:SP6
vector_side:right
27938..81075
misc_feature

```

```
misc_feature      /note="assembly_fragment"
81176..179947
/note="assembly_fragment"
BASE COUNT      59341 a 31579 c 31342 g 57385 t 300 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 69; DB 61; Length 179947;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17382 ttgttgcagctgagtgcaatggctgatctcagctcactgcaacctcgctccgcg 17441
|||||
Db 12883 TTGTTGCCAGGCTGAGTGCAATGGCGTGATCTCAGCTCACTGCAACCTCGCCTCCG 12942

Qy 17442 ggttcaagc 17450
|||||
Db 12943 GGTTCAGC 12951

RESULT 93
AL355305 185257 bp DNA PRI 16-OCT-2000
LOCUS      Human DNA sequence from clone RP11-487F23 on chromosome 6, complete
DEFINITION
ACCESSION  AL355305
VERSION    AL355305.9 GI:10798327
KEYWORDS  HTG.
SOURCE     human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 185257)
Frankland,J.
Direct Submission
Submitted (16-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 11, 2000 this sequence version replaced gi:10715951.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-487F23 is from the library RPCI-11.2 constructed at the
Roswell Park Cancer Institute by the group of Pieter de Jong. For
further details see http://bacpac.med.buffalo.edu/
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-487F23 The true
left end of clone RP3-491C16 is at 115276 in this sequence. The
true right end of clone RP11-787I22 is at 59873 in this sequence.
FEATURES
Location/Qualifiers
1..185257
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-487F23"
/clone_lib="RPCI-11.2"
6..477
/note="match: GSS: Em:AQ702346"
7..453
/note="match: GSS: Em:AQ727088"
25..383
/note="match: GSS: Em:AQ714702"
2036..2336
/note="AluX repeat: matches 1..295 of consensus"
2653..3012
/note="180 copies 2 mer aa 56% conserved"
2787..2924
/note="46 copies 3 mer aga 68% conserved"
4483..4776
/note="AluY repeat: matches 1..296 of consensus"
4844..5161
/note="AluJo repeat: matches 1..312 of consensus"
5357..7857
/note="L1M4 repeat: matches 2318..4838 of consensus"
7858..8166
/note="AluJo repeat: matches 1..312 of consensus"
8167..8394
/note="L1M4 repeat: matches 4838..5063 of consensus"
8437..8566
/note="L2 repeat: matches 1676..1802 of consensus"
8567..8860
/note="AluSp repeat: matches 1..295 of consensus"
8861..9718
/note="L2 repeat: matches 1802..2710 of consensus"
9926..10773
/note="SVA repeat: matches 519..1358 of consensus"
10043..11362
/note="CpG island"
/evidence=not_experimental
10774..10877
/note="SVA repeat: matches 630..729 of consensus"
10878..11942
/note="SVA repeat: matches 1..954 of consensus"
10879
/note="Random repeat. Forced join. Gap size estimated to
be approximately 750bp by restriction enzyme digest data."
complement(11329..11551)
/note="Sequence from overlapping clones BA787I22 (EMBL
accession AL390208) & BA689K9 (EMBL accession AL357046).
Assembly confirmed by restriction digest."
12969..13498
/note="L1MA5A repeat: matches 5757..6295 of consensus"
complement(14960..15375)
/note="match: GSS: Em:AQ185137"
14970..14997
/note="14 copies 2 mer aa 89% conserved"
complement(15077..15388)
/note="match: GSS: Em:AQ036138"
complement(15185..15387)
/note="match: GSS: Em:AQ000515"
15490..15841
/note="MT1A1 repeat: matches 1..365 of consensus"
15857..15902
/note="MADE1 repeat: matches 1..46 of consensus"
15958..16082
/note="L2 repeat: matches 2626..2750 of consensus"
16651..16819
/note="L2 repeat: matches 2526..2703 of consensus"
16872..16903
/note="16 copies 2 mer tg 93% conserved"
17022..17176
/note="MIR repeat: matches 31..192 of consensus"
17204..17521
/note="L2 repeat: matches 2177..2481 of consensus"
complement(17772..18252)
/note="match: GSS: Em:AQ387274"
19206..19515
/note="AluSg1 repeat: matches 1..308 of consensus"
```

repeat_region 19952..20250
/note="AluJo repeat: matches 3. .304 of consensus"
repeat_region 20452..20513
/note="MTLJ repeat: matches 112. .181 of consensus"
repeat_region 20569..21116
/note="MTLJ repeat: matches 1. .531 of consensus"
repeat_region 21225..21341
/note="MTLJ repeat: matches 388. .516 of consensus"
misc_feature 21614..21689
/note="Sequence from overlapping clones ba787122 (EMBL accession AL390208). Assembly confirmed by restriction digest."
misc_feature 21690..21693
/note="Sequence from overlapping clone ba787122 (EMBL accession AL390208). Assembly confirmed by restriction digest."
repeat_region 21784..21922
/note="MER7A repeat: matches 1. .151 of consensus"
repeat_region 21923..22122
/note="AluJb repeat: matches 87. .287 of consensus"
repeat_region 22123..22364
/note="MER7A repeat: matches 151. .346 of consensus"
repeat_region 22335..22349
/note="U6 repeat: matches 1. .106 of consensus"
repeat_region 22889..22966
/note="39 copies 2 mer ca 70% conserved"
repeat_region 22548..22638
/note="AluX repeat: matches 2. .292 of consensus"
repeat_region 22639..22760
/note="L2 repeat: matches 2621. .2749 of consensus"
repeat_region 22857..226013
/note="AluJb repeat: matches 137. .291 of consensus"
repeat_region 226016..22624
/note="L2 repeat: matches 2236. .2487 of consensus"
repeat_region 22681..226794
/note="MER4A repeat: matches 1. .333 of consensus"
repeat_region 22718..22757
/note="20 copies 2 mer aa 85% conserved"
repeat_region 22837..22854
/note="89 copies 2 mer gg 65% conserved"
repeat_region 22622..22088
/note="LRK3 repeat: matches 1. .521 of consensus"
repeat_region 22132..22270
/note="HAL1 repeat: matches 155. .298 of consensus"
repeat_region 22547..22846
/note="AluSq repeat: matches 1. .299 of consensus"
repeat_region 22891..223085
/note="LIMC4 repeat: matches 7742. .7960 of consensus"
repeat_region 22143..22362
/note="HAL1 repeat: matches 625. .847 of consensus"
repeat_region 22373..223682
/note="AluJo repeat: matches 1. .294 of consensus"
repeat_region 223769..223887
/note="L2 repeat: matches 2003. .2130 of consensus"
repeat_region 22485..224595
/note="L2 repeat: matches 2134. .2245 of consensus"
repeat_region 22461..224738
/note="MER5A repeat: matches 9. .105 of consensus"
repeat_region 224920..225146
/note="L2 repeat: matches 2498. .2750 of consensus"
repeat_region 22523..225368
/note="MIR repeat: matches 23. .149 of consensus"
repeat_region 22684..227003
/note="MIR repeat: matches 2. .144 of consensus"
repeat_region 227053..227173
/note="L2 repeat: matches 2626. .2748 of consensus"
repeat_region 22792..228105
/note="MER7A repeat: matches 1. .346 of consensus"
misc_feature complement(38225..38666)
/note="Sequence from overlapping clones ba787122 (EMBL accession AL390208) & ba689K9 (EMBL accession AL357046). Assembly confirmed by restriction digest."
misc_feature complement(38668..38740)

/note="Sequence from overlapping clones ba787122 (EMBL accession AL390208) & ba689K9 (EMBL accession AL357046). Assembly confirmed by restriction digest."
Query Match 0.3%; Score 69; DB 90; Length 185257;
Best Local Similarity 100.0%; Pred. No. 5.1e-26;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8902 gtgcagtcgacacatcgcctcactcagcctccacccctccaggttcagattctcc 8961
|||||
Db 100632 GTGCAGTGGCACAATCTCGGCTCACTCGAGCTCCACCTCCCGAGTTCAAGTATTCTCC 100691
|||||
QY 8962 tgctctcagc 8970
|||||
Db 100692 TGCCTCAGC 100700
|||||
RESULT 94
AC046171
LOCUS
DEFINITION Homo sapiens chromosome 17 clone RP11-948G15 map 17, WORKING DRAFT
SEQUENCE, 14 unordered pieces.
AC046171
ACCESSION AC046171.3 GI:13184209
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 189149)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-948G15
JOURNAL Unpublished
RECORD 2 (bases 1 to 189149)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienna,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliou,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
TITLE
JOURNAL
COMMENT Submitted (13-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2001 this sequence version replaced gi:7712190.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7946
Center clone name: 948_G15
----- Summary Statistics
Sequencing vector: M13; M77815; 40% of reads
Sequencing vector: Plasmid; n/a; 60% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 185448 bases at least Q40
 Consensus quality: 186995 bases at least Q30
 Consensus quality: 187536 bases at least Q20
 Insert size: 190000; agarose-fp
 Insert size: 187849; sum-of-contigs
 Quality coverage: 10.0 in Q20 bases; agarose-fp
 Quality coverage: 10.1 in Q20.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 11116: contig of 11116 bp in length
 * 11117 11216: gap of 100 bp
 * 11217 13252: contig of 2036 bp in length
 * 13253 13352: gap of 100 bp
 * 13353 16535: contig of 3183 bp in length
 * 16536 16635: gap of 100 bp
 * 16636 20967: contig of 4332 bp in length
 * 20968 21067: gap of 100 bp
 * 21068 24711: contig of 3644 bp in length
 * 24712 24811: gap of 100 bp
 * 24812 52351: contig of 27540 bp in length
 * 52352 52451: gap of 100 bp
 * 52452 62074: contig of 9623 bp in length
 * 62075 62174: gap of 100 bp
 * 62175 70797: contig of 8623 bp in length
 * 70798 70897: gap of 100 bp
 * 70898 92634: contig of 21737 bp in length
 * 92635 92734: gap of 100 bp
 * 92735 106944: contig of 14210 bp in length
 * 106945 107044: gap of 100 bp
 * 107045 129146: contig of 22102 bp in length
 * 129147 129246: gap of 100 bp
 * 129247 157583: contig of 28337 bp in length
 * 157584 157683: gap of 100 bp
 * 157684 183142: contig of 25459 bp in length
 * 183143 183242: gap of 100 bp
 * 183243 189149: contig of 5907 bp in length.

FEATURES

Source

1..189149
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /clone_lib="RPCI-11 Human Male BAC"
 1..11116
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 11217..13252
 /note="assembly_fragment"
 13353..16535
 /note="assembly_fragment"
 16636..20967
 /note="assembly_fragment"
 21068..24711
 /note="assembly_fragment"
 24812..52351
 /note="assembly_fragment"
 52452..62074
 /note="assembly_fragment"
 62175..70797
 /note="assembly_fragment"
 70898..92634
 /note="assembly_fragment"
 92735..106944

misc_feature
 /note="assembly_fragment"
 107045..129146
 /note="assembly_fragment"
 129247..157583
 /note="assembly_fragment"
 157684..183142
 /note="assembly_fragment"
 183243..189149
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right"

BASE COUNT 50053 a 45663 c 44537 g 47593 t 1303 others
 ORIGIN

Query Match 0.3%; Score 69; DB 72; Length 189149;
 Best Local Similarity 100.0%; Pred. No. 5.2e-26;
 Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17555 caggctggtctgaactcctgacctgacctgagtgatccaccacacctcagctcccaaaagtgt 17614
 |||||
 DB 19041 CAGGCTGGTCTGAACTCTGACCTCAGGTGATCCACCCACCTCAGCCCTCCCAAGTGT 19100

QY 17615 gggattaca 17623
 |||||
 DB 19101 GGGATTACA 19109

RESULT 95

AC055866

LOCUS AC055866 204685 bp DNA HTG 10-MAR-2001

DEFINITION Homo sapiens chromosome 17 clone RP11-376M2 map 17, *** SEQUENCING
 IN PROGRESS ***, 9 unordered pieces.

ACCESSION AC055866.12 GI:13270714

VERSION HTG; HTGS_PHASE1.

KEYWORDS SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 204685)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 17, clone RP11-376M2

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 204685)

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
 Meuldorf,J., Meneus,L., Mihova,T., Miranda,C., Mieng,A., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigillo,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 10, 2001 this sequence version replaced gi:13249494.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE

JOURNAL

COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9812
Center clone name: 376_M_2

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 11849: contig of 11849 bp in length
* 11850 11949: gap of 100 bp
* 11950 14698: contig of 2749 bp in length
* 14699 14798: gap of 100 bp
* 14799 62250: contig of 47452 bp in length
* 62251 62350: gap of 100 bp
* 62351 101354: contig of 39004 bp in length
* 101355 101454: gap of 100 bp
* 101455 141387: contig of 39933 bp in length
* 141388 141487: gap of 100 bp
* 141488 145481: contig of 3994 bp in length
* 145482 145581: gap of 100 bp
* 145582 148600: contig of 3019 bp in length
* 148601 148700: gap of 100 bp
* 148701 157288: contig of 8588 bp in length
* 157289 157388: gap of 100 bp
* 157389 204685: contig of 47297 bp in length.

FEATURES
Source
1. .204685
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-376M2"
/clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 52778 a 48916 c 48886 g 53074 t 1031 others
ORIGIN

Query Match 0.3%; Score 69; DB 72; Length 204685;
Best Local Similarity 100.0%; Pred. No. 5.2e-26;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17555 caggctgggtcgaactccctcagctcaggtgatccaccctcagctcccaagtgtt 17614
|||||
Db 151649 CAGCGTGGTCTCGAATCTCTGACTCAGGTGATCCACCCACCTCAGCTCCCAAGTGT 151708
|||||
QY 17615 gggattaca 17623
|||||
Db 151709 GGGATTACA 151717

RESULT 96
AC027808
LOCUS AC027808 222779 bp DNA HTG 10-JAN-2001
DEFINITION Homo sapiens chromosome 15 clone RP11-351M8 map 15, WORKING DRAFT
SEQUENCE 3 unordered pieces.
ACCESSION AC027808
VERSION AC027808.3 GI:12061537
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 222779)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens chromosome 15, clone RP11-351M8
2 (bases 1 to 222779)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehocsky,J.,
Levine,R., Lieu,G., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 10, 2001 this sequence version replaced gi:7658437.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9092
Center clone name: 351_M_8
----- Summary Statistics
Sequencing vector: M13; M77815; 26% of reads
Sequencing method: Plasmid; n/a; 74% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 221849 bases at least Q40
Consensus quality: 222140 bases at least Q30
Consensus quality: 222333 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 222579; sum-of-contigs
Quality coverage: 18.7 in Q20 bases; agarose-fp
Quality coverage: 15.8 in Q20.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 15998: contig of 15998 bp in length
* 15999 16098: gap of 100 bp
* 16099 208293: contig of 192195 bp in length
* 208294 208393: gap of 100 bp
* 208394 222779: contig of 14386 bp in length.

FEATURES
Source
1. .222779
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-351M8"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature 1. .15998

HS64K7/c	HS64K7	139378 bp	DNA	PRI	19-FEB-2001
LOCUS	Human DNA sequence from clone RP1-64K7 on chromosome 20q11.21-11.22				
DEFINITION	Contains the EIF2S2 gene for eukaryotic translation initiation factor 2 subunit 2 (beta, 38kD), a putative novel gene, the gene for heterogeneous nuclear ribonucleoprotein RALY or autoantigen p542, an RPS2 (RPS4) (40S ribosomal protein S2) pseudogene, ESTs, STS, GSSs and two CpG islands, complete sequence.				
ACCESSION	AL031668	23	GI:11497492		
VERSION	AL031668.2				
KEYWORDS	HTG; CpG island; EIF2S2; RALY; RPS2; RPS4.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Laird,G.				
TITLE	Direct Submission				
JOURNAL	Submitted (19-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk				
COMMENT	requests: clonerequest@sanger.ac.uk On Dec 1, 2000 this sequence version replaced gi:10198603. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr20 IMPORTANT: This sequence is not the entire insert of clone RP1-64K7. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP1-64K7 is at 139378 in this sequence. The true left end of clone RP4-785619 is at 121863 in this sequence. The true right end of clone RP5-1125A11 is at 97 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP1-64K7 is from the library RPC1-1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pCYPAC2.				
FEATURES	Location/Qualifiers				
source	1..139378 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosomes="20" /map="q11.21-11.23" /clone="RP1-64K7" /clone_lib="RPC1-1" complement(1..324) /note="match: GSS: Em:AQ5444763" 280..369 /note="MER5B repeat: matches 87..177 of consensus" 325..779 /note="match: GSS: Em:AQ165395" 342..611 /note="match: GSS: Em:AQ564778" 344..899 /note="match: GSS: Em:AQ627096" complement(1230..1735)				
misc_feature					
repeat_region					
misc_feature					
misc_feature					
misc_feature					
misc_feature					

/note="assembly_fragment					
clone_end:SP6					
vector_side:left"					
16099..208293					
/note="assembly_fragment"					
208394..222779					
/note="assembly_fragment					
clone_end:T7					
vector_side:right"					
BASE COUNT	61389 a	53087 c	49603 g	58498 t	202 others
ORIGIN					
Query Match	0.3%; Score 69; DB 71; Length 222779;				
Best Local Similarity	100.0%; Pred. No. 5.3e-26;				
Matches	69; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY 17555	cagctgtctcgaactcagctcaggtgatccaccacccagctcccaagtgtt	17614			
Db 70011	CAGCTGCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCAGGCTCCCAAGGT	70070			
QY 17615	gggattaca	17623			
Db 70071	GGGATTACA	70079			
RESULT 97					
AF282036/c	644 bp	DNA	PRI	23-AUG-2000	
LOCUS	Human.				
DEFINITION	Homo sapiens clone 15qtel_c320at3 sequence.				
ACCESSION	AF282036				
VERSION	AF282036.1	GI:9885538			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 644)				
AUTHORS	Riethman,H.C. and Moyzis,R.K.				
TITLE	Integration of telomeric DNA sequences with the human reference sequence				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 644)				
AUTHORS	Riethman,H.C. and Moyzis,R.K.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-JUN-2000) Molecular Genetics, The Wistar Institute, 3601 Spruce St., Philadelphia, PA 19104, USA				
FEATURES	Location/Qualifiers				
source	1..644 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosomes="15" /map="15qtel" /clone="15qtel_c320at3" /note="cosmid end sequence from half-YAC 2183" 167 a 162 c 179 g 134 t 2 others				
BASE COUNT	167 a	162 c	179 g	134 t	2 others
ORIGIN					
Query Match	0.3%; Score 68; DB 89; Length 644;				
Best Local Similarity	100.0%; Pred. No. 8.6e-26;				
Matches	68; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY 17559	ctggtctcgaactcagctcaggtgatccaccacccagctcccaagtgttggga	17618			
Db 88	CTGCTCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCAGCTCCCAAGTGTGGGA	29			
QY 17619	ttacaggc	17626			
Db 28	TTACAGGC	21			
RESULT 98					


```

SEQUENCE, 5 unordered pieces.
AC020656
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152714)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbaria, J.,
Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowle, S., Brieva, M., Brown, M., Brown, N.P., Bryant, N.P., Buhay, C.,
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flag, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C.,
Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Louisege, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, M., Mawhney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
Ogulu, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoostari, N.,
Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wlecyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 152714)
Worley, K.C.
Direct Submission
Submitted (08-JAN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 28, 2001 this sequence version replaced gi:12965309.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMfF
Center clone name: RP11-1143G9
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodypy; 20% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 148489 bases at least Q40
Consensus quality: 152411 bases at least Q30
Consensus quality: 154053 bases at least Q20

Estimated insert size: 149939; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 5 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1. 97950: contig of 97950 bp in length
* 97951 98050: gap of unknown length
* 98051 148979: contig of 50929 bp in length
* 148980 150219: gap of unknown length
* 150220 150319: contig of 1140 bp in length
* 150320 151444: gap of unknown length
* 151445 151544: contig of 1125 bp in length
* 151545 152714: gap of unknown length
* 152714: contig of 1170 bp in length.
Location/Qualifiers
1. 152714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="12"
/clone="RP11-1143G9"
BASE COUNT 46678 a 30161 c 29004 g 46466 t 405 others
ORIGIN

Query Match 0.3%; Score 68; DB 66; Length 152714;
Best Local Similarity 100.0%; Pred.No. 1.8e-25;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18973 gagatcgagaccatcctggctaaacatgatgaaccccgctctactaaaaatacaaaa
|||||
Db 149341 GAGATCGAGACCATCTGGCTACATGATGAACCCGCTCTACTAAAAATACAAAAT 149400

QY 19033 tagctggg 19040
|||||
Db 149401 TAGCTGGG 149408

RESULT 100
AL359085
LOCUS
DEFINITION Human DNA sequence from clone RP11-113J24 on chromosome 13,
complete sequence.
ACCESSION AL359085
VERSION AL359085.14 GI:12956942
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152794)
Pearce, A.
Direct Submission
Submitted (22-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Feb 16, 2001 this sequence version replaced gi:12830334.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate

```

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; SW.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13> RP11-113J24 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

This sequence is the entire insert of clone RP11-113J24 The true left end of clone RP11-153023 is at 145239 in this sequence. The true right end of clone RP11-237L12 is at 41597 in this sequence. Location/Qualifiers

```

source
1..152794
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-113J24"
/clone_lib="RPCI-11.1"
807..1096
/notes="AluSg repeat: matches 3. .292 of consensus"
1495..2031
/notes="CpG island"
/evidence=not_experimental
2881..3186
/notes="AluYa5 repeat: matches 1. .307 of consensus"
3361..3659
/notes="AluJb repeat: matches 1. .297 of consensus"
3778..3962
/notes="MER5A repeat: matches 2. .186 of consensus"
3984..4051
/notes="MER5A repeat: matches 123. .189 of consensus"
4481..4869
/notes="MSTA repeat: matches 1. .426 of consensus"
5322..5578
/notes="MER58B repeat: matches 20. .336 of consensus"
6616..6777
/notes="MIR repeat: matches 19. .192 of consensus"
7017..7182
/notes="MER90 repeat: matches 435. .603 of consensus"
7917..8094
/notes="LTR28 repeat: matches 1. .178 of consensus"
8121..8192
/notes="LTR1 repeat: matches 119. .190 of consensus"
8149..8435
/notes="LTR27 repeat: matches 289. .580 of consensus"
9063..9343
/notes="AluJo repeat: matches 3. .289 of consensus"
9420..9566
/notes="MIR repeat: matches 93. .249 of consensus"
9567..9796
/notes="MLT1C repeat: matches 225. .464 of consensus"
10281..10503
/notes="MLT1H repeat: matches 49. .295 of consensus"
10703..10776
/notes="MLT2CB repeat: matches 1. .70 of consensus"
10777..11069
/notes="AluSx repeat: matches 1. .293 of consensus"
11079..11120
/notes="MLT1F repeat: matches 469. .514 of consensus"
11093..11125
/notes="MLT1F repeat: matches 481. .512 of consensus"
11139..11438
/notes="AluSx repeat: matches 7. .306 of consensus"

repeat_region
11893..12012
/notes="LMC5 repeat: matches 7670. .7794 of consensus"
12241..12541
/notes="AluSx repeat: matches 1. .302 of consensus"
12980..13289
/notes="MER33 repeat: matches 1. .324 of consensus"
13355..13648
/notes="AluSx repeat: matches 1. .312 of consensus"
13847..14185
/notes="L2 repeat: matches 1821. .2177 of consensus"
14285..14343
/notes="L2 repeat: matches 2336. .2399 of consensus"
14351..14543
/notes="MER20 repeat: matches 6. .213 of consensus"
14657..14808
/notes="L2 repeat: matches 2579. .2747 of consensus"
14916..15050
/notes="MIR repeat: matches 23. .169 of consensus"
15152..15461
/notes="AluJb repeat: matches 1. .312 of consensus"
16383..16681
/notes="AluSx repeat: matches 1. .299 of consensus"
20295..20591
/notes="AluSx repeat: matches 2. .298 of consensus"
22740..22783
/notes="11 copies 4 mer acac 81% conserved"
22856..22911
/notes="MIR repeat: matches 98. .153 of consensus"
24363..24668
/notes="AluY repeat: matches 1. .306 of consensus"
24675..24799
/notes="MER5A repeat: matches 61. .186 of consensus"
25712..25753
/notes="21 copies 2 mer ac 97% conserved"
25714..25753
/notes="10 copies 4 mer acac 100% conserved"
26029..26422
/notes="L1M4A repeat: matches 5872. .6294 of consensus"
26818..26905
/notes="44 copies 2 mer cc 67% conserved"
26826..26905
/notes="20 copies 4 mer tccc 81% conserved"
26907..27197
/notes="AluSp repeat: matches 1. .292 of consensus"
27557..27610
/notes="27 copies 2 mer aa 77% conserved"
27631..27785
/notes="MIR repeat: matches 52. .216 of consensus"
28187..28498
/notes="AluSg repeat: matches 2. .308 of consensus"
28510..28811
/notes="AluSx repeat: matches 1. .286 of consensus"
28812..29271
/notes="L1PA16 repeat: matches 5715. .6156 of consensus"
29272..29559
/notes="AluSg1 repeat: matches 1. .288 of consensus"
29560..30186
/notes="L1PA16 repeat: matches 5113. .5715 of consensus"
30187..30494
/notes="AluJb repeat: matches 1. .296 of consensus"
30495..31503
/notes="L1PA16 repeat: matches 4045. .5113 of consensus"
31551..32301
/notes="SVA repeat: matches 4. .748 of consensus"
32385..32432
/notes="12 copies 4 mer caca 97% conserved"
33888..33919
/notes="8 copies 4 mer aaac 93% conserved"
34478..34783
/notes="AluY repeat: matches 1. .306 of consensus"
35441..35802
/notes="THE1B repeat: matches 1. .364 of consensus"
35845..35910
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2001, 20:35:50 ; Search time 18106.1 Seconds
(without alignments)
13920.749 Million cell updates/sec

Title: US-09-434-382-28
Perfect score: 26664
Sequence: 1 tatcagggtgactgaattcta.....ttcgcaaggtctttttgaca 26664

Scoring table: OLIGO.NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 8
Total number of hits satisfying chosen parameters: 20453455

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estin5.*
68: em_estom1.*
69: em_estom2.*
70: em_estov1.*
71: em_estov2.*
72: em_estpl1.*
73: em_estpl2.*
74: em_estpl3.*
75: em_estpl4.*
76: em_estpl5.*
77: em_estpl6.*
78: em_estpl7.*
79: em_estpl8.*
80: em_estpl9.*
81: em_estpl10.*
82: em_estro1.*
83: em_estro2.*
84: em_estro3.*
85: em_estro4.*
86: em_estro5.*
87: em_estro6.*
88: em_estro7.*
89: em_estro8.*
90: em_estro9.*
91: em_estro10.*
92: em_estro11.*
93: em_estro12.*
94: em_estro13.*
95: em_estro14.*
96: em_estro15.*
97: em_estro16.*
98: em_estro17.*
99: em_estro18.*
100: em_estro19.*
101: em_estro20.*
102: gb_est25.*
103: gb_est26.*
104: gb_est27.*
105: gb_est28.*
106: gb_est29.*
107: gb_est30.*
108: gb_est31.*
109: gb_est32.*
110: gb_est33.*
111: gb_est34.*
112: gb_est35.*
113: gb_est36.*
114: gb_est37.*
115: gb_est38.*
116: gb_est39.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Score	Query Match	Length	DB ID	Description				
1	521	2.0	523	144	BF116134	7n78a06.x	198	0.7	416
2	518	1.9	518	144	BF115696	7n78a09.x	198	0.7	650
3	509	1.9	975	165	BE250412	600943455	195	0.7	115
4	478	1.8	531	118	AW572950	hfl17h05.x	193	0.7	414
5	448	1.7	499	117	AW510825	hfl17h05.x	193	0.7	552
6	435	1.6	456	7	AA410756	z39b08.r	190	0.7	860
7	429	1.6	478	118	AW592601	hf45a09.x	182	0.7	424
8	425	1.6	527	16	AI089646	qb16g07.x	180	0.7	482
9	414	1.6	525	225	AQ218458	HS-3247.B	179	0.7	577
10	413	1.5	478	171	BF963225	QV2-NN004	178	0.7	398
11	409	1.5	506	171	BF959904	QV2-NN004	173	0.6	406
12	408	1.5	492	256	B44326	HS-1059-A1-	171	0.6	318
13	398	1.5	673	153	BG421943	602449690	171	0.6	222
14	381	1.4	511	225	AQ175652	HS-3213.B	169	0.6	220
15	380	1.4	627	167	BE386924	601274815	164	0.6	447
16	351	1.3	440	7	AA410664	zt29g11.r	162	0.6	372
17	343	1.3	396	12	AA811170	OD42603.s	161	0.6	374
18	338	1.3	345	115	AW407520	UI-HF-BM0	158	0.6	421
19	328	1.2	475	147	BF364978	CM2-NN114	158	0.6	451
20	325	1.2	531	103	AI937465	wp77e01.x	158	0.6	526
21	321	1.2	584	107	AU127299	AI127299	157	0.6	228
22	319	1.2	696	108	AU138595	AU138595	157	0.6	317
23	302	1.1	491	10	AA676661	z167h01.s	157	0.6	493
24	298	1.1	514	171	BF954354	QV2-NN004	157	0.6	553
25	295	1.1	346	169	BF769772	IL5-IT002	156	0.6	354
26	292	1.1	343	170	BF825929	CM4-HN002	155	0.6	726
27	290	1.1	494	141	BE858252	7g21a09.x	153	0.6	370
28	284	1.1	394	113	AA248468	2820640.3	153	0.6	677
29	283	1.1	489	102	AI803400	tc42f03.x	152	0.6	754
30	281	1.1	489	8	AA534478	nf76f10.s	144	0.5	157
31	277	1.0	670	108	AU143668	AU143668	142	0.5	390
32	276	1.0	746	153	BG395840	602458622	141	0.5	278
33	275	1.0	461	10	AA635046	ab48b06.r	141	0.5	278
34	274	1.0	666	108	AU141334	AU141334	141	0.5	278
35	273	1.0	533	169	BF816722	MR2-CI012	141	0.5	278
36	269	1.0	742	175	BG253351	602362946	141	0.5	278
37	268	1.0	442	105	AL039527	DKF2p434B	141	0.5	278
38	268	1.0	712	107	AU126037	AU126037	141	0.5	278
39	263	1.0	727	174	BG166177	602340918	141	0.5	278
40	262	1.0	657	108	AU138795	AU138795	141	0.5	278
41	261	1.0	356	20	AI475638	tc86d11.x	141	0.5	278
42	261	1.0	457	147	BF309436	601892128	141	0.5	278
43	255	1.0	410	15	AI033108	ow98g08.s	141	0.5	278
44	255	1.0	461	118	AW592223	hf41a01.x	141	0.5	278
45	254	1.0	359	147	BF371674	BF371674	141	0.5	278
46	253	0.9	479	11	AA716607	z968g07.s	141	0.5	278
47	245	0.9	429	12	AA838624	oe91f04.s	141	0.5	278
48	236	0.9	691	167	BE409312	601300940	141	0.5	278
49	235	0.9	439	17	AI201492	qs74b03.x	141	0.5	278
50	234	0.9	947	147	BF305817	601889183	141	0.5	278
51	232	0.9	865	8	AA522537	nl38e08.s	141	0.5	278
52	228	0.9	282	13	AA928608	om75b03.s	141	0.5	278
53	227	0.9	433	19	AI357786	qu98d07.x	141	0.5	278
54	222	0.8	416	171	BF944518	CM1-NN021	141	0.5	278
55	221	0.8	249	113	AW247380	2820640.5	141	0.5	278
56	220	0.8	477	102	AI804749	tu42d02.x	141	0.5	278
57	214	0.8	316	7	AA464307	zx78c04.r	141	0.5	278
58	214	0.8	316	117	AW511765	xu76f03.x	141	0.5	278
59	214	0.8	334	171	BF914840	IL3-UT011	141	0.5	278
60	212	0.8	437	145	BF203824	601868811	141	0.5	278
61	212	0.8	452	152	BG327066	602426274	141	0.5	278
62	212	0.8	535	108	AU158921	AU158921	141	0.5	278
63	210	0.8	220	227	AQ349651	RPC111-12	141	0.5	278
64	209	0.8	225	112	AW175582	QV0-BT004	141	0.5	278
65	207	0.8	219	8	AA504146	aa59e06.s	141	0.5	278
66	205	0.8	409	149	BF477438	nac60h05.	141	0.5	278
67	204	0.8	541	108	AU148489	AU148489	141	0.5	278

ALIGNMENTS

RESULT 1

BF116134

LOCUS

7n78a06.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3570706 3', mRNA EST 24-OCT-2000

DEFINITION

7n78a06.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3570706 3', mRNA EST

ACCESSION

BF116134.1

VERSION

BF116134.1

KEYWORDS

GI:10985610

SOURCE

EST.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 523)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 479.
Location/Qualifiers
1. 523
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3570706"
/clone_lib="NCI_CGAP_Ov18"

/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(df) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCGACATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 81 a 160 c 167 g 115 t
ORIGIN

Query Match 2.0%; Score 521; DB 144; Length 523;
Best Local Similarity 100.0%; Pred. No. 7e-165;
Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1017 gctcgccacctgcgcagcagagaagcgcgacccgtcggtgtctccgcgcccaaa 1076
Db 1 GCTCGCGCACCTGCGCACGAGAGAGCGCGACCGTGGGGTCTCCGGCGGCCAAA 60
QY 1077 caactgtacctcaggtggtgagcgggttagccgggactcggcgcggtctactgt 1136
Db 61 CACCCTGTACCTGCAAGTGGTGGCAGCGGTAGCGGGACTCGGGCGCGCTCTAGCT 120
QY 1137 ctctccagttcaaccgtcagttcaacgacgacgccccgtcccgctggccctcagt 1196
Db 121 CTCTCCGAGTTCAACCGGTCTAGTCAACGAGCCACGCCCGTCCCGCTGGGCCCTCAGT 180
QY 1197 gggcgagcctctgagatcggggacacctccagggttcggcttccctgctcacat 1256
Db 181 CGCGCAGCCTCTGAGATCGGGGACCTCCAGGGCTTCGGCTTCCTGCTCACAT 240
QY 1257 gtggttoactgttcgggggttcgtgaggttatgtggtggtggaatccgagattcttg 1316
Db 241 GTGTTTCTACTGTTCGGGGGTTCGTGGAGTTATGTTGGTGGTGGAAATCCGAGATTTT 300
QY 1317 catccatgtattctcgagatctggaacttcagaccttcggagctccacctcttcgttc 1376
Db 301 CATCCATGTGATTTCTCGGAGTCTGTGAAGAACTTCAGGCTTCGGGTCTGAGGCTCCTTT 360
QY 1377 cccaaaccttgcccgccgtggtgctgacgaccttcggagctccacctcttcctgctgc 1436
Db 361 CCCAACCTTGCGCCCGCGCTGGCTGTGACGACTTTCGAGCTCCACCTCTCCGTGC 420
QY 1437 acccaaggccagtgctgctgttagcgtgtggtggtggaacagatctcgtgtgtagccgg 1496
Db 421 ACCCAAGGCCAGTGTGCTGTGTGTAGCGTGTGGGTGGACAGATCTGCTGTAGCCGG 480
QY 1497 tgggtgagaaggactatttctgctagcacccacaca 1537
Db 481 TGGTGGAGAAAGGACTCATTTTGTCTAGCACCCACACA 521

RESULT 2
LOCUS BF115696 518 bp mRNA EST 24-OCT-2000
DEFINITION 7n82g09.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3571384 3', mRNA sequence.
ACCESSION BF115696
VERSION BF115696.1 GI:10985172
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 518)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL • Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 493.
Location/Qualifiers
1. 518
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3571384"
/lab_host="NCI_CGAP_Ov18"
/tissue_type="fibrotheoma"
/note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(df) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCGACATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 76 a 159 c 172 g 111 t
ORIGIN

Query Match 1.9%; Score 518; DB 144; Length 518;
Best Local Similarity 100.0%; Pred. No. 7.2e-164;
Matches 518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 999 gcgcgccgagcaccgctgcgcacccgtcgacacgtgtacctgcagtggtgagcggtagcgggactc 1058
Db 1 GCGCGCGCGAAGGACCGCGTGGCGACCTGGCACCGGAGAGAGCGCGGACCGTGGG 60
QY 1059 gtgtcccgcgccgccccaaacacggtgtacctgcagtggtgagcggtagcgggactc 1118
Db 61 GTGCTCCGGCGGCCAAAACACCGTACCTGCAGGTGGTGGCAGCGGTAGCGGGACTC 120
QY 1119 gggcgccgctcagtcgtctctccagagttcaaccggtcagttcaacgacgacgacgcccgt 1178
Db 121 GGGCGCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTCAACGAGCCACGCCCGGT 180
QY 1179 cccgctggccctcagtgcgcgacgctctgagcatcggggacacctccagggcttcg 1238
Db 181 CCGCTGGGCCCTCAGTGGCGGACGCTCTGAGGATCGGGGACCTCCAGGGCTTCGG 240
QY 1239 ctccctgcttcacacatgtgttccactgtgtcgggggttcgtgaggtatgtgtggtggtg 1298
Db 241 CTTCCTGCTTACACATGTGTTTCTGCGGGGTTCGTGAGTTATGTTGGTGGTGG 300
QY 1299 gaaatccagattcttctcatccatgtattctcgtcgtatctgtgaaagaacttaaggct 1358
Db 301 GAAATCCAGAGATCTTTCATCCATGTGATTTCTCGGATCTGTGAAGAATCTCAGGCT 360
QY 1359 gggctctgagctcctttcccaaccttggcccgccgctggtcgtcagcacttccgag 1418
Db 361 GGGTCTGAGGCTCTTTTCCCAACCTTGGGCCCGCGCTGGCTGTGAGACTTTCGAG 420
QY 1419 ctccacctcttcggtgcaccccaaggccagtcgtgtcgtttagctgtgtggtggtgaca 1478
Db 421 CTCCACCTCTTCCGTGACCCCAAGGCCAGTGTGCTGTTGTAGCTGTGGGGTGGACA 480
QY 1479 gatctggtgttagcgggtggtgagaaagactcatt 1516
Db 481 CATGCTGTGTAGCCGCTGGTGGAGAAAGGACTCAT 518


```
RESULT 3
LOCUS BE250412/c 975 bp mRNA EST 13-JUL-2000
DEFINITION 600943455T1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960077 3',
mRNA sequence.
ACCESSION BE250412
VERSION BE250412.1 GI:9120523
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 975)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM52 row: n column: 14
High quality sequence start: 22
High quality sequence stop: 756.
FEATURES
source
1. 975
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2960077"
/clone_lib="NIH_MGC_17"
/tissue_type="rhodomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pORF7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```
BASE COUNT 185 a 296 c 258 g 236 t
ORIGIN
```

```
Query Match 1.9%; Score 509; DB 165; Length 975;
Best Local Similarity 100.0%; Pred. No. 6.1e-161;
Matches 509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 25810 aggtctgttgagactttccaacaatgccaaagctgattccccactgaagccctgt 25869
|||||
```

```
Db 606 AGGCTGCTTTGGAGACTTTCACAAATGCCCAAGCTGATTCCTCCCTCAAGAGCCCTGT 547
|||||
```

```
Qy 25870 ttgctggcagacatcagagagatgagagagcaggggaaacggagctgcgaggtgc 25929
|||||
```

```
Db 546 TTGCTGGCGGACATCGAGGAGATGGAGGAGCGCAGGAGAAAGCGGAGCTGCGGAGGTGC 487
|||||
```

```
Qy 25930 gggcgccctcctgtccagggagctggcagcgccctggagatgggagcctcagcaga 25989
|||||
```

```
Db 486 GGGGGGCCCTCTGTCCAGGAGAGTGGCAGCGCGCTGGAGGATGGGGAGCCCTCAGCAGA 427
|||||
```

```
Qy 25990 agcggggccacacagagagccacagggccaagaaggttcagagccagtggaatctggga 26049
|||||
```

```
Db 426 AGCGGGCCACACAGAGAGCCACAGGCCAAGAAGGTGAGAGCCAGCCAGTGAAGATCGGA 367
|||||
```

```
Qy 26050 gacctgaactcaagagctgtgtcttctgtcccccagcagcccgatctgcctc 26109
|||||
```

```
Db 366 GACCTTGAACCTCAGAAGGCTGTGTCTTCTGCGCCACGACGACCGGTATCTGCGCCTC 307
|||||
```

```
Qy 26110 cttgctgtagaagctgaagagcagcgtgtccccaggagcagctcaggataggtgtatg 26169
|||||
```

```
Db 306 CTTGCTGGTAGAAGCTGAAGAGCACGGTCCCCAGGAGGAGCTCAGGATAGGTGGTATG 247
|||||
Qy 26170 gagctgcccagagcttgggtcccccataaagaactagctctatagatgcctcttagact 26229
|||||
Db 246 GAGCTGTGCCGAGAGCTTTGGGCTCCACATAAGCACTAGTCTATAGATGCCCTCTTAGGACT 187
|||||
Qy 26230 ggtcctggcacagccgcccagggaggtgccacacgggaagcaagcagatgaactaat 26289
|||||
Db 186 GGTCCCTGGCACACGCCGGCGGCGGAGGCTGCCACACGGGAAGCAAGCAGATGAACATAAT 127
|||||
Qy 26290 ttcatttcaagcaggttttttaaaagatc 26318
|||||
Db 126 TTCATTTCAAGGCAGTGTTTTAAAGAAGTC 98
|||||
```

```
RESULT 4
```

```
AW572950/c
```

```
LOCUS
```

```
DEFINITION
```

```
AW572950 531 bp mRNA EST 13-MAR-2000
```

```
IMAGE:2932185 3', mRNA sequence.
```

```
ACCESSION
```

```
AW572950
```

```
VERSION
```

```
AW572950.1 GI:7237683
```

```
KEYWORDS
```

```
EST.
```

```
SOURCE
```

```
human.
```

```
ORGANISM
```

```
Homo sapiens
```

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE
```

```
1 (bases 1 to 531)
```

```
AUTHORS
```

```
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

```
TITLE
```

```
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
```

```
Tumor Gene Index
```

```
Unpublished (1997)
```

```
CONTACT: Robert Strausberg, Ph.D.
```

```
Email: cgabbs-r@mail.nih.gov
```

```
This clone is available royalty-free through LLNL; contact the
```

```
IMAGE Consortium (info@image.llnl.gov) for further information.
```

```
Seq primer: -40UP from Gibco
```

```
High quality sequence stop: 459.
```

```
FEATURES
```

```
source
```

```
1. 531
```

```
Location/Qualifiers
```

```
/organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
```

```
/clone="IMAGE:2932185"
```

```
/clone_lib="Soares_NFL_T_GBC_S1"
```

```
/lab_host="DH10B"
```

```
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
```

```
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
```

```
Equal amounts of plasmid DNA from three normalized
```

```
libraries (fetal lung NbHL19W, testis NHT, and B-cell
```

```
NCI-CGAP-CCB1) were mixed, and ss circles were made in
```

```
vitro. Following HAP purification, this DNA was used as
```

```
tracer in a subtractive hybridization reaction. The driver
```

```
was PCR-amplified cDNAs from pools of 5,000 clones made
```

```
from the same 3 libraries. The pools consisted of
```

```
I.M.A.G.E. clones 297480-302087, 682632-687239,
```

```
72408-728711, and 729096-731399. Subtraction by Bento
```

```
Soares and M. Fatima Bonaldo.
```

```
BASE COUNT 99 a 154 c 136 g 142 t
```

```
ORIGIN
```

```
Query Match 1.8%; Score 478; DB 118; Length 531;
```

```
Best Local Similarity 99.8%; Pred. No. 2e-150;
```

```
Matches 528; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 25938 ctctgtccaggagctggcagcgccctgagatgggagcctcagcagaagcgccc 25997
|||||
```

```
Db 531 CTCCTGTCAGGGAGCTGGCGAGCGGCTCGAGGATGGGAGCCTCAGCAGACGGGGCC 472
|||||
```

```
Qy 25998 cacacagagagccacagggccaagaagtcagagccctcagagctgaagatctgggagacctga 26057
|||||
```

```
Db 471 CACACAGAGGAGCCACAGGCCAAGAGGTCAGAGCCCGCAGAGTGAAGATCTGGGAGAGCCCTGA 412
|||||
```



```
FEATURES
source
Location/Qualifiers
1..456
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:755991"
/clone_lib="Soares ovary tumor NBHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/notes="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACATCTGAATGGGAGCGCGCGGTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library constructed by Bento, Soares and
M.Fatima Bonaldo."
BASE COUNT 147 a 95 c 94 g 120 t
ORIGIN

Query Match 1.6%; Score 439; DB 7; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.8e-137;
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13995 qcactctgccttaagtcactgttgtagttggtggctccggtgtacacagcctcaag 14054
|||||
DB 456 GCACCTCTGCTTAAGTCACCTGTGGTAGTGGTGGGCTCCGGTGTACAGCCCTCAAG 397
|||||

QY 14055 tgaattagaaaagattgaaaactagaaaactgagactagaaattcaactagaactc 14114
|||||
DB 396 TGAAATTAGAAAAGATTGAAAACCTAGAAACACTGAGGACTAGAAATTCACCTAGACTC 337
|||||

QY 14115 ttacagctctataccagaagaattctagaactttttgaattcttaactaagcccccag 14174
|||||
DB 336 TTACAGCTCTTATACCAAGAAATTTAGAACTTTTGTAAATCTTAATGCCCCAG 277
|||||

QY 14175 attatcatttggattatttgaactgaattatttttccattacactgaatgaacaa 14234
|||||
DB 276 ATTATCAATTTGGATATTTTGAACCTGAATTAATTTTCTCCATPACCTGATGAAACAA 217
|||||

QY 14235 atgagtggtgcagagtgtgagactgtcgtgtgcaagatccgtgttatggatggac 14294
|||||
DB 216 ATGAGTGGGTGCAGAGTGTGACACTGTCGTGTCAGAGTCCGTGTTATGGATGGAC 157
|||||

QY 14295 tcacagctggggaatgtcttttgggttaactgccactctgtgtgttcctctatcgaagt 14354
|||||
DB 156 TCACAGCTGGGAATGTCTTTGGGCTAACTGCCACTCTGTTGTTGCTCTATCGAAGT 97
|||||

QY 14355 taaccagtttgcggttcagctttcattccagatggaaatcatctttgacccacctatctg 14414
|||||
DB 96 TAACCACTTTTGGGTTTCAGCTTTTCATTCAGATGGAATCATCTTTTACCCACCTATCTG 37
|||||

QY 14415 agtttgaatcttttcccc 14433
|||||
DB 36 AGTTTGAATCTTTCCCCC 18
|||||

RESULT 7
AW592601/c 478 bp mRNA EST 22-MAR-2000
LOCUS h45409.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2934808 3', mRNA sequence.
ACCESSION AW592601
VERSION AW592601.1 GI:7279786
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 478)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps@email.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Gibco
High quality sequence stop: 451.
Location/Qualifiers
1..478
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2934808"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 92 a 132 c 122 g 132 t
ORIGIN

Query Match 1.6%; Score 425; DB 118; Length 478;
Best Local Similarity 99.8%; Pred. No. 1.4e-132;
Matches 475; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 25991 qcgggccacacagagagccacagagcccaagaagtgtagccagccagatctcggag 26050
|||||
DB 478 GCGGGCCACACAGAGGAGCCACAGGCCNAGAGGTGAGAGCCAGTGAAGATCTGGGAG 419
|||||
QY 26051 accctgaactcagaagctgtgtctcttgcacacagcagccagctatctgcctcc 26110
|||||
DB 418 ACCCTGAACCTCAGAAGGCTGTGTCTTCTGCCCCACGACGACCCGTATCTGCCCTCC 359
|||||
QY 26111 ttgctgtagaagctgaagacagcagcgggtcccccagagagcagctcagtaggtatcg 26170
|||||
DB 358 TTGCTGTAGAAGCTGAAGACACGCTCCCCAGGAGCAGCTCAGGATAGTGGTATGG 299
|||||
QY 26171 agctgtccgaggtctgggtcccccacacataagcactagtctatagatgcctcttagactg 26230
|||||
DB 298 AGCTGTGCCGAGGCTTGGGCTCCCCACATNAGCACTAGTCTATAGATGCTCTTAGGACTG 239
|||||
QY 26231 gtgctgtgacagcgcgcgggcccagagggtgccacacaggaagcagcagatgaactaatt 26290
|||||
DB 238 GTGCTGSCACAGCCGCGGCGCAGAGGCTGCCACACGGAAGCAGAGAGTAAT 179
|||||
QY 26291 tcattcaaggcagtttttaaaagactcttggaacacagcggcgacaccttctctaat 26350
|||||
DB 178 TCATTTCAAGGCAGTTTTTAAAGAGTCAATGAAACAGCGCGCACCTTTCTCTTAAT 119
|||||
QY 26351 ccagaaagtatccctgcacacacagacagcagagatcagtggtctctaa 26410
|||||
DB 118 CCAGCAAGTGTATTCCTTGCACACACAGACAGACAGAGTAAACAGGATCAGTGGGCTCAA 59
|||||
QY 26411 gttcccgagacttaacgaaaaatagttatttcagctgcaataaagattgagttgcaa 26466
|||||
DB 58 GTGCTCCGAGACTTAACGAAATAGTATTTTACAGTGCAATAAAGATTGAGTTGGCAA 3
|||||

RESULT 8
AI089646/c 527 bp mRNA EST 18-AUG-1998
LOCUS qbl6g07.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
DEFINITION


```
QY 13943 aagagctgtgttttgggaagaccactatctggttttacaggttcagagcgccactcct 14002
|||||
Db 254 AAGAGCTGTGTGTTGGGAAACACCATATCTGGGTTTACAGTTTCAGAGGCGGCACCTCT 195
|||||
QY 14003 gccttaagtcactgttgtagtgggtccggttcacacagcctcaagtgaaatta 14062
|||||
Db 194 GCCTTAAGTCACGTGTTGGTGGGCTCCGGTTCACACAGCCCAAGTGAAATTA 135
|||||
QY 14063 gaaaagattgaaactgaagcaactgagactagagaattcaactagaactctt 14116
|||||
Db 134 GAAAAGATTGAARACTAGAACAACTCAGGACTAGAAATTCAACTAGAACTCTT 81
|||||

RESULT 10
BF963225 478 bp mRNA EST 22-JAN-2001
LOCUS QV2-NN0045-201200-574-d07 NN0045 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF963225
ACCESSION BF963225.1 GI:12380500
VERSION BF963225.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 478)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV2&t2=QV2-NN0045-
201200-574-d07&t3=2000-12-20&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 478.

FEATURES
source
1..478
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN0045"
/dev_stage="Adult"
/note="Organ: nervous_normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 141 a 87 c 131 g 119 t
ORIGIN

Query Match 1.5%; Score 413; DB 171; Length 478;
Best Local Similarity 99.8%; Pred. No. 1.6e-128;
Matches 463; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5950 gtaataacccttttgcacagacatgaagaggtttgcacagatagatttttaa 6009
|||||
Db 15 GTAATACCTCTTTTGTATCAGACGATCATGAAGAGGTTTCACAGATAGATTTTTTAAA 74
|||||
```

```
QY 6010 taaataatgattacagcaacctaagaaagtgtgttgggttttagaagctcctgcaaat 6069
|||||
Db 75 TAATAATGATTACAGCAACCTAAAAGAAAGTGTGTGGGGTTAGAACCTCTCGCAAT 134
|||||
QY 6070 tccgaagtatcagggccagatgtagtggcttagcttaggaaaagagttagcttcttgcct 6129
|||||
Db 135 TCCGAAGTATCAGGCCAGATGATGTGCTCTAGCTTAGGAAAAGAGTTAGTCTTGTCTCT 194
|||||
QY 6130 tgaacttggtctaaagacattcattctggttttacttactatgtgaagagagatccaagca 6189
|||||
Db 195 TGAACCTGGCTAAAGACATTCATGCTGTGTTTACTTACATGTGAAGAGATACCAAGCA 254
|||||
QY 6190 gtagggttattctctgttagtactaactaatgtgtagcttactaaagttagtctgtaggg 6249
|||||
Db 255 GTAGGGGTATTCTCTGTGTACTACTAATGATGCTTACCAAGTAGTGTGATGGG 314
|||||
QY 6250 tgacagaccagagcaccagcaagggccagagaagtagtcagacacctggcagagagatgagg 6309
|||||
Db 315 TGACAGACCAGAGCACCCAGCAAGAGGCCAGAGAAAGTCCAGAACCTGGCGAGAGATGAGG 374
|||||
QY 6310 ctacactgactgaagcagagaagcagcagcagcagcagcagcagcagcagcagcagcagc 6369
|||||
Db 375 CTTACTACTGACTGAAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 434
|||||
QY 6370 caagtctctagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6413
|||||
Db 435 CAAGTCTCTTAGGCCAGTGTGCTGTGATGAGCTGATCAGCACTCC 478
|||||
```

RESULT 11

```
BF959904
LOCUS BF959904 506 bp mRNA EST 22-JAN-2001
DEFINITION QV2-NN0045-051200-516-d05 NN0045 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF959904
VERSION BF959904.1 GI:12377179
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 506)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV2&t2=QV2-NN0045-
051200-516-d05&t3=2000-12-05&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 504.

FEATURES
source
1..506
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN0045"
/dev_stage="Adult"
/note="Organ: nervous_normal; Vector: puc18; Site 1: SmaI;
```


Best Local Similarity 100.0%; Pred. No. 3e-103;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26127 aagagcaggtccccagagagcagctcaggtatagtgatgagctgtgccgaggtctt 26186
|||||
Db 345 AAGACACGGTCCCCCAGGAGGACCTCAGGATAGTGTGTGGAGCTGTGCCGAGGCTT 286
QY 26187 gggctccacataaagcactagctatagatgcctcttaggactgtgctgtggcagcagcg 26246
|||||
Db 285 GGGCTCCACATAAGCACTAGTCTATAGATGCTCTTAGGAGTGTGGCTGGCAGCGG 226
QY 26247 cggccagaggtgtccacaggaagcagatgaactaatcttcaaggcagtt 26306
|||||
Db 225 CGGGCAGGAGGCTCCACAGGAAGCAAGCAGATCAACTTAATTTCAATTTCAAGGCAGTT 166
QY 26307 tttaagaagtcttgaacacagcggcgccactttctcttaataccagcaaatgattcc 26366
|||||
Db 165 TTTAAAGAAGTCTTGGAAACAGACGGCGGCACTTTCTCTTAATCCAGCAAGTATTCC 106
QY 26367 ctgcacacagagacaagcagagtaacaggtacagtgctggtctaaagtgtccgagacttaac 26426
|||||
Db 105 CTGCACACAGAGACAAGCAGAGTAACAGGATCAGTGGGTCTAAGTGTCCGAGACTTAAC 46
QY 26427 gaaaatagttattcagctgcataaagattgagttgc 26464
|||||
Db 45 GAAATAGTATTTCAGCTGCAATAAAGATTGAGTTTGC 8

RESULT 19
BF364978/c
LOCUS BF364978 475 bp mRNA EST 24-NOV-2000
DEFINITION CM2-NN1145-230900-384-508 NN1145 Homo sapiens cdna, mRNA sequence.
ACCESSION BF364978
VERSION BF364978.1 GI:11327003
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 475)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&t2=CM2-NN1145-
230900-384-b08&t3=2000-09-23&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 473.
Location/Qualifiers
1. 475
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN1145"
/dev_stage="Adult"

FEATURES source
1. 475
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN1145"
/dev_stage="Adult"
/note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions.*
BASE COUNT 78 a 162 c 126 g 109 t
ORIGIN

Query Match 1.2%; Score 328; DB 147; Length 475;
Best Local Similarity 99.5%; Pred. No. 6.2e-100;
Matches 428; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 25858 tgaagccctgtttctggtcgacatcgagagatgagagcgagcgaggaagcgaggagc 25917
|||||
Db 475 TGAAGCCCTGTTTCTGCGGNACATCGAGGAGATCGAGGAGCGCAGGAGAGCGGGAGC 416
QY 25918 tgcgcaggtgcggcgccctctctgtccagggagctggcagggcgccctggaggtgggg 25977
|||||
Db 415 TCGGCAGGTGCGGCGGCCCTCTCTGTCCAGGAGCTGCGAGGCGCCTGGAGGATGGGG 356
QY 25978 agctcagcagaagcgggcgccacacagagggagcagagcaggaaggtcagagccagt 26037
|||||
Db 355 AGCCTCAGCAAGAGCGGGCCACACAGAGAGGCCACAGGCCAAGAGGTCAGAGGCCAGT 296
QY 26038 gaagatctgggagaccctgaactcagaagctgtgtctctgtccacgcagcagcacc 26097
|||||
Db 295 GAAGATCTGGGAGAGACCTTGAACTCAGAGGCTGTGTCTCTCTGCCCCACGACGCCACCC 236
QY 26098 gtatcgccctctctgtggttagaagctgaagagcacggtgcccccagggagggcagctcagg 26157
|||||
Db 235 GTATCTGCCCTCTTGTGTTGGTAGAAGCTGAAGAGCACAGGTTCCCCAGGAGGAGCTCAGG 176
QY 26158 ataggtgtatggagctgtgcgaggttgggtcccccataaagcactagctctatagtg 26217
|||||
Db 175 ATAGTGTATGGAGCTGTGCCGAGGCTTGGCTCCCATTAAGACACTAGTCTATAGATG 116
QY 26218 cctcttagactgtgctgtgcacagcgcgggccagagagctgtgccacaggaagcagc 26277
|||||
Db 115 CCTTTAGACTGTGCTGTGCACAGCGCGGGCCAGGAGGCTGCCACACGGAAGCAAGC 56
QY 26278 agatgaacta 26287
|||||
Db 55 AGATGAACCTA 46

RESULT 20
AI937465/c
LOCUS AI937465 531 bp mRNA EST 08-MAR-2000
DEFINITION wp77e01.x1 NCI_CGAP_Brn25 Homo sapiens cdna clone IMAGE:2467800 3',
mRNA sequence.
ACCESSION AI937465
VERSION AI937465.1 GI:5676335
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 531)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGA), Tumor Gene Index
Unpublished (1998)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 1291 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 467.

FEATURES

BASE COUNT
ORIGIN

Query Match 1.2%; Score 325; DB 103; Length 531;
Best Local Similarity 99.2%; Pred. No. 6.1e-99;
Matches 525; Conservative 0; Mismatches 4; Indels 0

QY	25938	ctcctgtccaggagctggcaggcgccctggagatgggagacctcagcagaagcgggcc	25997
Db	531	CTCCTGTCCAGGAGCTGGCAGCGCGCTGGAGATGGGAGCCTCAGCAGAAGCGGGCC	472
QY	25998	cacacagagagccacagagcccaagaaggtcagagcccagtgaaagatctgtggaagaccctga	26057
Db	471	CACACAGAGAGGACCACAGGCCAAGAAGGTcAGAGCCcAGTGAAGATCTGGGAGACCCTGA	412
QY	26058	actcagaaggtgtgtgtctcttgcgccacgcagcagcaccgtatctgcctctcttctgtg	26117
Db	411	ACTCAGAAGGCTGTGTCTCTTCTGCCCCACGCACGCCACCCGTATCTGCCCTCTCTTGTCTGG	352
QY	26118	tagaagctgaagcagcaggtgtcccccagagagcgagctcagataggtgagttagagctgtg	26177
Db	351	TAGAAGCTGAAGAGCAGCGTCCCCACGAGGAGCAGCTCAGGATAGGTGATGGAGCTGTG	292
QY	26178	ccagagcttgggtcccaataagcactagtctatagatgcctctttaggactggtgcctg	26237
Db	291	CCGAGGGTTGGGTGCCACATAAGCACTAGTCTATAGATGCCTTTAGGACTGGTGCTGTG	232
QY	26238	gcacagcgcggccagaggaaggtgtgcacacaggaagcagatgaactaatctcatttc	26297
Db	231	GCACAGCTGCGGGCCAGGAGGCTGCCACACGGAAAGCAGATGAATTAATTCATTTTC	172
QY	26298	aaggcagtttttaagaagctcttggaaacagacgcgccaccitctcctaactccagcaa	26357
Db	171	AAGCAGTTTTTTAAAGAAGTCATGGAACACAGCGCGGCACCTTTCTCTTAATCCACGAA	112
QY	26358	agtgattccctgcacaccagagacaagcagtagtaaacaggatcagtggtcttaagtctcg	26417
Db	111	AATGATTCCCTGCACACCAGAGACAACACAGATTAACAGGATCAGTGGGTCTAAGTGTCCG	52
QY	26418	agacttaacgaaatagtatatttcagctgcaataaagattgagtttgcga	26466
Db	51	AGACTTAACGAAATAGTATTTTTCAGCTGCAATTAAGGATTTGAGTTTTCGAA	3

RESULT	21
AU127299	
LOCUS	
DEFINITION	
ACCESSION	

RESULT	22
AU138595	
LOCUS	696 bp mRNA
DEFINITION	PLACE1 Homo sapiens cDNA clone PLACE1008903 5', mRNA
ACCESSION	AU138595
VERSION	AU138595.1 GI:110000116

AUL27299.1 GI:10952015
 EST.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 584)
 Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
 Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
 Isogai,T.
 HRI human cDNA project
 Unpublished (2000)
 Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.

BASE COUNT	122 a	166 c	178 g	115 t	3 others
ORIGIN	cells after 2-weeks retinoic acid (RA) induction				

	Query Match	1.2%; Score 321; DB 107; Length 584;	
	Best Local Similarity	100.0%; Pred. No. 1.3e-97;	
	Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	836	acgggtcagttgtagcgaggaaacgcggcgctaggtaccgcgcgcttctcagitttg	895
Db	1	AGCGGCTAGGTGGTGCACGGGAACGGGGCGTAGTGACCGCGGCTTTCTCAGITTTG	60
Qy	896	gtggagacggcgcatgtggcgctttgtcgcgtgctgcggtccggcgcgacgcacca	955
Db	61	GTGAGACGGCGCATGTGGCGCTTTGCTTCGCTGCTGCCGCGCCGACGCACCA	120
Qy	956	tgtcgaggacgcacaatctgcaggacccccgccgcgcgcgcgcgcgaagacc	1015
Db	121	TGTCGAGGGACGCACCATTATCGAGGCACCCGCCGCCGAGCGCGCAAGSACC	180
Qy	1016	cgrtcggcgacctgcgcacgcgagagaagacgcgcacgtcgagggtactccggcgcgcccaa	1075
Db	181	CGETCGGGCACCTCGCCACCGGAGAAAGCGCGACCGCTCGGGGTGCTCCGGCGGCCCAA	240
Qy	1076	acaccgttgtacctcgagttggtggcacgvggtagcgcgggactcgggcgcgcgctctacy	1135
Db	241	ACACCGTGTACCTGCAGTGGTGGCAGCGGTACCGGGACTCGGGGCGCGCTCTACG	300
Qy	1136	tcttctccgagttcaaacccgtt	1156
Db	301	TCTTCTCCGAGTTCAACCGGT	321

RESULT	22
AUI38595	
LOCUS	AUI38595 696 bp mRNA
DEFINITION	AUI38595 PLACE1 Homo sapiens cDNA clone PLACE1008903 5', mRNA EST 25-OCT-2000
ACCESSION	AUI38595
VERSION	AUI38595.1 GI:11000116


```
QY 26430 aatagtttcagctgcaataaagattgagttgc 26464
|||||
Db 35 AATAGTATTTCACTGCAATTAAGATTGAGTTTGC 1

RESULT 24
BF954354 514 bp mRNA EST 22-JAN-2001
LOCUS QV2-NN0045-131100-414-e12 NN0045 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF954354
ACCESSION BF954354
VERSION BF954354.1 GI:12371629
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 514)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV2&t2=QV2-NN0045-
131100-414-e12&t3=2000-11-13&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 476.

FEATURES
Location/Qualifiers
1..514
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN0045"
/dev_stage="Adult"
/note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
143 a 104 c 138 g 129 t

BASE COUNT 143 a 104 c 138 g 129 t
ORIGIN

Query Match 1.1%; Score 298; DB 171; Length 514;
Best Local Similarity 99.7%; Pred. No. 7.4e-90;
Matches 348; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6082 gggccagatgctggtcttagcttagaagaagttagttcttgccttgaaacttgcta 6141
|||||
Db 131 GGGCCAGATGATGCTGCTTAGCTTAGAAGAGATTAGTCTTCTCTGAACCTGGCTA 190

QY 6142 aagacattcgtctgttttttactatcatgtgaagagagaccagcagtaggggtattt 6201
|||||
Db 191 AAGACATTCATGCTGCTTTTACTATGATGTGAAGAGAGTACCAAGCAGTAGGGGTATTT 250

QY 6202 ccttggttagtactaactaattgtatgcttactaagttagtctgatgggtgacagaccaga 6261
|||||
```


T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pMT3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo.
BASE COUNT 95 a 140 c 124 g 135 t
ORIGIN

Query Match 1.1%; Score 290; DB 141; Length 494;
Best Local Similarity 99.2%; Pred. No. 3.7e-87;
Matches 490; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 25971 gatgggagcctcagcagaagcgggccacacagagagcagcagcgaagaggtcaga 26030
Db 494 GATGGGAGCCTCAGCAGAGCGGGCCACACAGAGGAGCCACAGGCCAAGAGGTGAGA 435
QY 26031 gccagtgaaagctctgggagaccctgaactcagaaggtgtgtcttctgccccacgca 26090
Db 434 GCCCAGTGAAGATCTGGGAGACCTGAACCTCAGAAGGCTGTGTCTCTCTGCCCCAGGCA 375
QY 26091 cgcaccgtatctgcctctctgtgtagaagctgaagacagcggtccccaggagga 26150
Db 374 CGCACCGGTATCTGCCCTCCTTGTGTGAAGCTGAAGAGCAGCGGTCCCCCAGGAGGCA 315
QY 26151 gctcagataggtggtatgagctgtgccgaggtgtggctccacacataagcactagtct 26210
Db 314 GCTCAGATAGGTGATGAGAGCTGTGCCAGGCTTGGGGTCCCATATAGCACTAGTCT 255
QY 26211 atagatgcctcttaggactggtgctggcagacccgcgggcaggaggtgcccacacgga 26270
Db 254 ATAGATGCCCTTAGGACTGGTGGCAGACAGCTGGCGGCCAGGAGGCTGCCACACGGA 195
QY 26271 agcaagcagatgaactaatttcattcaaggcaggttttttaaagaagctcttggaacagac 26330
Db 194 AGCAAGCAGATGAACCTAATTTCAATTTCAAGGCAGTTTTTAAAGAAGTCATGGAACAGAC 135
QY 26331 ggcgcacctctctcctaaccagaaagtattccctgcacacagacagacagcagagt 26390
Db 134 GCGCGCACCTTCTCTTAATCCAGCAAAATGATTCCTGTGCACACAGACAGCAGAGT 75
QY 26391 aacaggatcagtggtgtaagtgtccgagacttaacgaaatagttattcagctgcaata 26450
Db 74 AACAGGATCAGTGGGTCTAAGTGTCCGAGACTTAACGAAAATAGTATTTTTCAGCTGCAATA 15
QY 26451 aagattgatttc 26464
Db 14 AAGATTGAGTTTC 1

RESULT 28
AW248468/c
LOCUS
DEFINITION 2820640.3prime NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2820640 3',
mRNA sequence.
ACCESSION AW248468
VERSION AW248468.1 GI:6591461
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 394)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Other ESTs: 2820640.5prime
Contact: Robert Strausberg, Ph.D.
Email: cgapps-f@mail.nih.gov
Tissue Procurement: DCTD/DTT cDNA Library Preparation: Ling
Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E.
Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing
project Clone distribution: MGC clone distribution information can

be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/brp/image/image.html Base Calling / Quality
Scores: PHRED from University of Washington Genome Center. Vector
Trimming: cross_match from University of Washington Genome Center
PHRAP suite. Poly-r Identification: patmatch.pl from Berkeley
Drosophila Genome Project. University of Washington Genome Center:
http://www.genome.washington.edu Polyadenylation: Based upon the
presence of a XhoI site followed by a run of 14 or more T residues
at the beginning of the sequence, this cDNA insert was
polyadenylated.
Plate: LCM4 row: L column: 17
High quality sequence stop: 213.
FEATURES
Location/Qualifiers
source 1..394
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2820640"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dr priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit.
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 80 a 103 c 97 g 114 t
ORIGIN
Query Match 1.1%; Score 284; DB 113; Length 394;
Best Local Similarity 99.5%; Pred. No. 4.2e-85;
Matches 384; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 26081 gccccacgcagcaccgctatctccctcctctgctgtagagctgagagcagcaggtccc 26140
Db 394 GCCCCACGCACGACCCGCTATCTGCCCTCTCTGCTGGTAGAAGCTGAAGACACAGCGTCCC 335
QY 26141 ccaggagcagctcagataggtggtgagagctgtgccgaggttgggtccccacataa 26200
Db 334 CCAGGAGCAGCTCAGGATAGGTGATGGAGCTGTGCCAGGCTTGGGCTCCACATAA 275
QY 26201 gcactagtctatagatgcctcttaggactggtgctggcagacgcgcggccaggaggtc 26260
Db 274 GCCTAGCTATAGATGCCTCTTAGGACTGGTGTGCTGTGCACAGCGCGGCGCAGAGGTT 215
QY 26261 gccacacggaagcagcagatgaactaatttcatttcaaggcagtttttaagaagcttt 26320
Db 214 GCCACAGGAAGCAAGCAGATGAACCTAATTTCAATTTCAAGGCAGTTTTTAAAGAAGTCAT 155
QY 26321 ggaacacagcggcgccaccttctcttaactcagcaagtgattccctgcacacagaga 26380
Db 154 GGAACACAGCGCGGCACCTTCTCTTAATCCAGCAAGTCATTCCTGTGCACACAGAGA 95
QY 26381 caagcagagtaacaggatcagtggtgtctaagtgtccgagacttaacgaaaaatagttttc 26440
Db 94 CAAGCAGAGTAACAGGATCAGTGGGTCTAAGTGTCCGAGACTTAACGAAAATAGTATTTC 35
QY 26441 agctgcataaagattgagtttgcga 26466
Db 34 AGCTGCAATAAAGATTGAGTTTGCAA 9
RESULT 29
AI803400/c
LOCUS
DEFINITION tc42f03.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
IMAGE:2067293 3', mRNA sequence.
ACCESSION AI803400
VERSION AI803400.1 GI:5368962


```
/clone="THYRO1000421"
/clone_lib="THYRO1"
/tissue_type="thyroid gland"
/note="Vector: pME18SFL3"
BASE COUNT 158 a 186 c 190 g 129 t 3 others
ORIGIN

Query Match 1.0%; Score 274; DB 108; Length 666;
Best Local Similarity 100.0%; Pred. No. 8.1e-82;
Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 883 ttctcagtttggtagagcgccgcatgtggcgcttggctgcgtgcgtgcgcgcg 942
Db 2 TTTCTCAGTTTGGTGGAGACGGCGCATGTGGCGCTTTGCTCGCTGCTGCGTCCGCG 61
QY 943 gccgagcgaccatgtgcagggagcgaccacatcgcaggcaccgcccccgcgcgcg 1002
Db 62 GCCGGACGCCACCATGTGCGAGGACGCCACCATATCGCAGGCACCGCCGCCGCGAGCGG 121
QY 1003 ccgcgaagaccgctgcgcgaccctgcgcagcgagagaagcgaccgctcggggtgc 1062
Db 122 CCCGCAAGACCCGCTCGCGACCTGCGCACGAGAGAGCGGACCGTCCGGGTGC 181
QY 1063 tccggcgcccaaacaccgtgtacctgcaggtgtgtgagcgaggtgagcgactcgggc 1122
Db 182 TCCGGCGGCCCAACACCGTGTACCTGCAGTGTGCGACGCGGTAGCGGGACTCGGGC 241
QY 1123 gccgcgctctactcttctccgagttcaaccggt 1156
Db 242 GCCGCGCTCTACGCTCTCTCCGAGTTCAACCGGT 275

RESULT 35
BF816722/c
LOCUS BF816722 533 bp mRNA EST 13-JAN-2001
DEFINITION MR2-C10128-071200-011-e04 C10128 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF816722
VERSION BF816722.1 GI:12152705
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 533)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=MR2&t2=MR2-C10128-
071200-011-e04&t3=2000-12-07&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 8
High quality sequence stop: 532.
Location/Qualifiers
1..533
/organism="Homo sapiens"

FEATURES
source
```

```
/db_xref="taxon:9606"
/clone_lib="C10128"
/dev_stage="Adult"
/note="Organ: colon_ins; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 114 a 149 c 145 g 125 t
ORIGIN

Query Match 1.0%; Score 273; DB 169; Length 533;
Best Local Similarity 99.2%; Pred. No. 1.9e-81;
Matches 473; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 21198 gacatcagaatatatacaatctgggggtgtgttctctgtgagtgagacatgcaataaagc 21257
Db 525 GACATCAGAATAATACAACTCTGGGTGTGTCTCTGTGGATGAGGACATGCATAAAGC 466
QY 21258 agcttgagtgagcggcctctcccggtgagatcctctgggggaagagggtttttg 21317
Db 465 AGCTTGGAGTGAGCGCGCTCTCCGGGGCTGAGATCCTGGGGAAAGAGGCTTTTGG 406
QY 21318 agttgacctgacacctgagcagcttttgaaccagctgaagctaataaggaggtgct 21377
Db 405 AGTTTGACCGACACACCTGCGAGCAGCTTTTGAACCAAGCTGAAGCTAATGGGAAGGTGCT 346
QY 21378 attgcaaccttgctccgctcccgactcttttcccccagaaggttaattcttagcac 21437
Db 345 ATTGCCACCTTGCTCTCGCTCCGACTCCGTGGTCCCCAGAGAGTAATGCTTAGCAC 286
QY 21438 cggggctctctctgcaaatgggtgcagccctcagtgctgctgctcctccagaga 21497
Db 285 CGGGGCTTCTCTCTGCAAAATGGTGCAGCCCTCTCAGTGTCTGCTGCTCCACAGAGA 226
QY 21498 atgaaggagccagagcgggtcagcactctctctgtgagcagcagcttctctgaaatgg 21557
Db 225 ATGAAGGAGGCCAGAGCGGGTCAAGCACTCTCTGCTTGGAGCAGAGCTTCTGAAATGG 166
QY 21558 actgcacagcagaatagcccaagaagttgtcagaatccagactccagagccctgcta 21617
Db 165 ACTGCACAGCAGAAATACCCCAAGAAAGTTGTCAAAATCCAGACTTCCAGAGCCCTGCCTA 106
QY 21618 aaaccaagtgcagaaaccccgagtcacacctggggagtcgtgcttaactgggtccctga 21674
Db 105 AAACCAAGTCAGAAACCCCGAGTGACACCTGGGAGTCTGCGTTAACTGGCTCCCTGA 49

RESULT 36
BG253351
LOCUS BG253351 742 bp mRNA EST 13-FEB-2001
DEFINITION G02362946F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4471427 5',
mRNA sequence.
ACCESSION BG253351
VERSION BG253351.1 GI:12763167
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 742)
NIH-MGC http://mnc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Prepared by: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LHAM10290 row: k column: 12
High quality sequence stop: 682.

FEATURES
source
1. .742
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4471427"
/clone_lib="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 191 a 194 c 221 g 136 t
ORIGIN

Query Match 1.0%; Score 269; DB 175; Length 742;
Best Local Similarity 99.4%; Pred. No. 3.7e-80;
Matches 489; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 25810 aggtctgttgagactttcaacaattgccaaagctattccccacactgaagccctgt 25869
|||||
Db 71 AGGCTGCTGTGGAGACTTTCACAAATGCCCAAGCTGATCCCCCATGAAGGCCCTGT 130
QY 25870 ttgctggcgacatcgaggagatggaggagcgcaggaggaagcggagctgcggcaggtgc 25929
|||||
Db 131 TTGCTGGCGACATCGAGGATGGAGAGCGCAGGAGGAAGCGGAGCTCGGCAGGTGC 190
QY 25930 gggcgccctcctgtccaggagctgycaggcgccctggagatggggagcctcagcaga 25989
|||||
Db 191 GGGCGGCCCTCTGTGCCA-GGAGCTGGCAGCGCGCTGGAGATGGGAGCCTCAGCAGA 249
QY 25990 aqcgggccacacagagagcgcacagggccaaagaaggtcagagccagtgagatcggga 26049
|||||
Db 250 AGCGGGCCACACAGAGGAGGCCAGGCCACAGAGAGGTGAGAGCCAGTGAAGATCTGGGA 309
QY 26050 gaccctgaactcagaagctgtgtcttctgccccacgcagcccgatctcgcctc 26109
|||||
Db 310 GACCTGAATCAGAAGCTGTGTCTCTGCCCCACGCGACCGTATCTGCCCTC 369
QY 26110 cttgctgtagaagctgaagagcagcgttccccagagcgagctcagataggtgtatg 26169
|||||
Db 370 CTTGCTGTAGAACTGAAGAGCAGCGGTCCCGCAGGAGCAGCTCAGGATAGGTGGTATG 429
QY 26170 gagctgtccaggtctgggtccacataagcactagtctataatgccctcttaggact 26229
|||||
Db 430 GAGCTGTCCGAGGCTTGGGTGCCACATAGCACTAGTCTATAGATGCCCTTTAGGACT 489
QY 26230 ggtgctggcacagcgcgycggccaggaggtgcccacacgcagcagcagcagatgaactaat 26289
|||||
Db 490 GGTGCTGGCAGAGCTGGCGGCCAGGAGGCTGCCACACGGAAGCAGCAGATGAAT 549
QY 26290 ttcattcaag 26301
|||||
Db 550 TTCATTTCAAGG 561

RESULT 37
AL039527 442 bp mRNA EST 29-FEB-2000
LOCUS
DEFINITION DKFZp434B2311.r1 434 (synonym: htes3) Homo sapiens cDNA clone
ACCESSION AL039527
VERSION AL039527.1 GI:5408569
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann ,S.
EST (Duesterhoeft, et al.)
Unpublished (1999)
Contact: Duesterhoeft A
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.
s1 sequence also available.
This clone (DKFZp434B2311) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
source
1. .442
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp434B2311"
/clone_lib="434 (synonym: htes3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT 142 a 62 c 63 g 170 t 5 others
ORIGIN

Query Match 1.0%; Score 268; DB 105; Length 442;
Best Local Similarity 99.7%; Pred. No. 9.7e-80;
Matches 318; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 9213 tgtttgtttctcaactccatccatcaggcctttattcaagtccttttaactctgtttact 9272
Db 1 TGTGTGTTTTCTTCACTCCATCCATCAGGCCTTATTCAAGCTTTTAAGCTCTGTTTACT 60
QY 9273 ttattcattccctgcgaatagctgaagctctaacaccagattaatgggaattagctag 9332
Db 61 TTATTTTCATCCCTGCAATAGCTAAGGCTAACACCAGATTAAATGGAATATTAGCTAG 120
QY 9333 cattcaaaagcctagatctgtaactctgaaattggcacaattccattcaaaattttg 9392
|||||
Db 121 CATTACAAAGGCCNAGATCTGTAACCTGAAATTTGGTCAAAATTCATTAATAATTTTG 180
QY 9393 ttacaataagctgtttgtaagatctgactagtggtctatttttaataagaattttgcatta 9452
Db 181 TTACAAATAGCTGTTTGTGAAGATCTGACTAGTGGCTTATTTTAATAGAAATTTGCATTA 240
QY 9453 aaatttatcaatacaatttgcacaaatttgtctaaatatgtgaaagatttcattgcc 9512
Db 241 AAATTTTATCAATACAAATTTGCAACAAATTTGCTCTAAATATGCTGAAAGATTTCATTGCC 300
QY 9513 tttttgtgggttagatta 9531
|||||
Db 301 TTTTGTGGCTTAGATTA 319

RESULT 38
AUI26037 712 bp mRNA EST 23-OCT-2000
LOCUS
DEFINITION AUI26037 NT2RM4 Homo sapiens cDNA clone NT2RM4002610 5', mRNA
sequence.
ACCESSION AUI26037
VERSION AUI26037.1 GI:10950753
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

LOCUS	AU138795	057 bp	mRNA	EST	200
DEFINITION	AU138795	PLACE1	Homo sapiens	cDNA clone PLACE1009311	5', mRNA
ACCESSION	AU138795				

IMAGE:2934408 3', mRNA sequence.

ACCESSION
AW592223
VERSION
AW592223.1 GI:7279399
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 461)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Seq primer: -400P from Gibco

High quality sequence stop: 450.

Location/Qualifiers

1..461

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="2934408"

/clone_lib="Soares_NFL_T_GBC_S1"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NBHL19W, testis NHT, and B-cell

NCI-CGAP_GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo. "

BASE COUNT 93 a 125 c 114 g 129 t

ORIGIN

Query Match 1.0%; Score 255; DB 118; Length 461;
Best Local Similarity 99.1%; Pred. No. 2.3e-75;
Matches 455; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 26008 agccacaggcccaagagtcagagcccgtagagatctgggagaccctgaacctcagaag 26067

|||||
Db 461 AGCCACAGGCCCAAGAGGTCAGAGGCCCGTAGAGATCTGGGAGACCCCTGAAGCTCAGAAGG 402

QY 26068 ctgtgtgtctctgtccacagcagcagcccgtagtctgcctccttgcgtggtagaagctga 26127

|||||
Db 401 CHTGTGTGCTTCTGCCCCAGCAGCACCCCGTATCGCCCTCTTGTGCTGTAGAGCTGA 342

QY 26128 agagcacagtgctcccccagaggagcagctcaggatagtggtgtatgagctgtgcgagcgcttg 26187

|||||
Db 341 AGAGCACAGTCCCCAGGAGGCGAGCTCAGGATAGTGTGTGAGCTGTGCCGAGGCTTG 282

QY 26188 ggctccacataagcactagtctatagatgcctcttagactggtgcctgacagccgc 26247

|||||
Db 281 GGTGCCCATGAAGACACATAGTCTATAGATGCTCTTAGACTGGTGGCTGCACAGCTGC 222

QY 26248 gggcagagagctgcacacggaagcaagacagatgaactaatttcatttcaggcagttt 26307

|||||
Db 221 GGGCCAGAGAGGCTGCCACAGGAAACAGACAGATGAATTTTCATTTCAAGGCAGTTT 162

QY 26308 ttaaagaagctcttgaacacagacggcgccaccccttcctcctaaccagcaaaagtgtattccc 26367

|||||
Db 161 TTAAAGAAGTCATGGAACACAGACGCGGACCTTTCCTTAATCCAGCAAAATGATTCCC 102

QY 26368 tgcacacagagacagaagcagagtaacaggaatgaatgggtcttaagtgtccgagactaacg 26427

|||||
Db 101 TGCACAGCAGACAAAGCAGAGTAACAGAGTACAGTGGTCTAAGTGTCCGAGACTTAACG 42

QY 26428 aaaaagtatttcagctgcaataaagattgagttgcaa 26466

|||||
Db 41 AAAATAGTATTTCAGCTGCATAAAGATTGAGTTTGCAA 3

RESULT 45

BF371674

LOCUS

BF371674

DEFINITION

RC5-FN0165-180700-011-C07 FN0165 Homo sapiens cDNA, mRNA sequence.

ACCESSION

BF371674

VERSION

BF371674.1 GI:11333699

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 359)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC6&t2=RC6-FN0165-180700-011-G07&t3=2000-07-18&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 51

High quality sequence stop: 358.

FEATURES

Location/Qualifiers

1..359

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="FN0165"

/dev_stage="Adult"

/note="Organ: prostate normal; Vector: puc18; Site_1: SmaI

; Site_2: SmaI; A mini-library was made by cloning

products derived from ORESIES PCR (U.S. Letters Patent

application No. 196,716 - Ludwig Institute for Cancer

Research) profiles into the puc 18 vector. Reverse

transcription of tissue mRNA and cDNA amplification were

performed under low stringency conditions."

BASE COUNT 82 a 66 c 99 g 112 t

ORIGIN

Query Match 1.0%; Score 254; DB 147; Length 359;

Best Local Similarity 99.7%; Pred. No. 5.4e-75;

Matches 304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 14578 tctgctttattctctgtggtgaatttgatgcgctgtagccccggctttgtatttaa 14637

|||||
Db 48 TCTGCTTTATTCTCTGTGGGTGAATTTGGATCGCGTGTGTAGCCCGGCTTTGATTATTAA 107

QY 14638 tccagttttggcagcaaaaacctcttcaatgaatcagtgctgattgagaccatgtgtg 14697

|||||
Db 108 TCCAGTTTGGCAGCAAAACCTGTTCAATGAATCAGGTGTTCATTTTGTAGAGCCATGTGTG 167

QY 14698 gatgtgtgatgtgctgggatagataaaaaatagctactgtgtatatatttttttaaagg 14757


```
|||||
Db 168 GATGTGTGATGATGCTGGGATAGATAAAATAGTACTGTATATATTTCTTTTAAAGGG 227
|||||
Qy 14758 aactggaggaaacacatcacatgcttagtaagtggtctctgttccagggtggaattt 14817
|||||
Db 228 AACTGGAGGAAACACATCAGATGTTAGTAAGTGGTCTGTGTGTCACAGGTGGTGAATTT 287
|||||
Qy 14818 cagatgattttcatttcgtgctgtctcagtgctcctctggaagcagacacacagggt 14877
|||||
Db 288 CAGATGATTTTCATTTCTGTCGCTGTGTCCTCAGTCTCTGGAAGCAGACACACAGGT 347
|||||
Qy 14878 ggcatt 14882
|||||
Db 348 GGCAT 352

RESULT 46
AA716607/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
AA716607 479 bp mRNA EST 29-DEC-1997
z968q07.s1 Soares_fetal_heart_NBHH19W Homo sapiens cDNA clone
IMAGE:398556 3', mRNA sequence.
AA716607
AA716607.1 GI:2728881
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 479)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marr,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 462.
Location/Qualifiers
1. .479
/organism="Homo sapiens"
/db_xref="GDB:1306379"
/db_xref="taxon:9606"
/clone="IMAGE:398556"
/clone_lib="Soares_fetal_heart_NBHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site:1: Not 1; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAATGGAGCGCCGATCTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."
93 a 127 c 118 g 141 t

BASE COUNT
ORIGIN

Query Match 0.9%; Score 253; DB 11; Length 479;
Best Local Similarity 99.1%; Pred. No. 1.1e-74;
```

```
Matches 453; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 26010 ccacagcccaagaaggtcagagccagtggaagatctggagaccctgaactcagaagct 26069
|||||
Db 467 CCACGCCCAAGAGGTCAGAGCCCAAGTGAAGATCTGGGAGACCTGNACTCAGAAGCT 408
|||||
Qy 26070 gtgtgtcttctgcccacacgcacgcccctctatctgcccctctctgtgtagaagctgaag 26129
|||||
Db 407 GTGTGTCTTCTGCCCCCAGCGACGCCGTATCTGCCCTCCTTCTGCTGTAGAGCTGAAG 348
|||||
Qy 26130 agcacggtccccccagggagcagtcaggtatggtggtatgagctgtgcccagccttggg 26189
|||||
Db 347 AGCACGGTCCCCCAGGAGGAGCTCAGGATAGGTGGTATGGAGGTGTCGCCAGGCTTGG 288
|||||
Qy 26190 ctcccacataagcactagtctatagatgcctcttagactgtgtgcctggcagacgcgcg 26249
|||||
Db 287 GTCCCATATAGCACTAGTCTATAGATGCCCTTATAGGACTGGTGGCCGACAGCTGGCG 228
|||||
Qy 26250 gccaggaggtgccacacggaagcaagcagatgaactaatttcatttcaggcgatgtttt 26309
|||||
Db 227 GCCAGGAGGTGCCACACGGAAGCAAGCAGATGAATTTTCATTTCAGGCGAGTTT 168
|||||
Qy 26310 aaagaagctcttggaaacagacgagcggcgacaccttctcttaaccagcaaatgattccctg 26369
|||||
Db 167 AAGAAGTTCATGGAACACAGACGGCGGACCTTCTCTTAATCCAGCAAAATGATTCCCTG 108
|||||
Qy 26370 cacaccagacacagagtaaacagatcagtggtgttaagttccagagacttaacgaa 26429
|||||
Db 107 CACACCAGAGACACAGCAGATGAAGATCAGTGGGTCTAAGTGTCCGAGACTTAACGAA 48
|||||
Qy 26430 aatagtatttcagctgcaataaagattgagttgtcaa 26466
|||||
Db 47 AATAGTATTTCAGTGCATTAATAAGATTGAGTTTGCAA 11
|||||

RESULT 47
AA838624/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
AA838624 429 bp mRNA EST 18-MAR-1998
ce91f04.s1 NCI_CGAP_Col2 Homo sapiens cDNA clone IMAGE:1419007,
mRNA sequence.
AA838624
AA838624.1 GI:2913423
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 429)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1787 Std Error: 0.00
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 375.
Location/Qualifiers
1. .429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1419007"
/clone_lib="NCI_CGAP_Col2"
/sex="mixed"
/tissue_type="colon tumor"

FEATURES
source
```



```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:979142"
/clone_lib="NCI_CGAP_Lu1"
/tissue_type="lung tumor"
/lab_host="SOLR (kanamycin resistant)"
/Note="Organ: lung; Vector: Bluescript SK-; Site:1: EcoRI;
Site:2: XhoI; Cloned unidirectionally. Primer: Oligo dT.
Bulk lung tumor. 5' adaptor sequence: 5' GAATTCGGCAGCAG 3'
3' adaptor sequence: 5' CTCAGTTTTTTTTTTTTTTT 3'
Average insert size: 1.1 kb."
BASE COUNT 179 a 235 c 227 g 218 t 6 others
ORIGIN

Query Match 0.9%; Score 232; DB 8; Length 865;
Best Local Similarity 99.5%; Pred. No. 1e-67;
Matches 402; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 26048 gagaccctgaactcagaaggtgtgtgtcttctgtcccccacacgaccccgatctgccc 26107
|
|
|
Db 403 GAGACCCTGAACCTCAGAAGGCTGTGTCTTCTGCCCCCAGCAGCAGCCCGTATCTGCC 344
|
|
|
Qy 26108 tccctgtcgttagaagctgaagagcaacggtcccccagagcagctcaggatagtgta 26167
|
|
|
Db 343 TCCTTGTCTGGTAGAAGCTGAAGAGCACGCTGCCCCAGGAGGAGCTCAGGATAGGTGTA 284
|
|
|
Qy 26168 tggagctgcgaggttggtctccacataagcaactagtctatagatcctcttagga 26227
|
|
|
Db 283 TGGAGCTGTGCGGAGGCTTGGGCTCCCACTAAGCACTAGTCTATAGATGCTCTTAGGA 224
|
|
|
Qy 26228 ctggtgcctggtgcagcgccgagcagagcgtgcacacgaaagcagcagataaacta 26287
|
|
|
Db 223 CTGTTGCTT-GCACAGCGCGGGCAGGAGGCTGCACACGAGCAGCAGTGAACCTA 165
|
|
|
Qy 26288 atttcattcagcagtttttaagaagcttcttgaaacagcagcgccacccctctct 26347
|
|
|
Db 164 ATTTTCATTCAAGCGAGTGTAAAGAAAGTCTATGAAACAGACGCGCGACCTTTCCTCT 105
|
|
|
Qy 26348 aatccagcaagtgatccctgcacaccagagacaagcagagtaacagatcagtgatc 26407
|
|
|
Db 104 AATCAGCAAAAGTATTCCTTCGACACACAGAGCAAGCAGAGTACAGATCAGTGGGTC 45
|
|
|
Qy 26408 taagtgtccgagacttaacgaaaatagttatttcagctgcaataa 26451
|
|
|
Db 44 TAAGTGTCCGAGACTTAACGAAAATAGTATTTCAGCTGCAATAA 1
```

```
RESULT 52
AA928608/c 282 bp mRNA EST 07-JUL-1998
LOCUS om75b03.sl NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1552973 3',
DEFINITION mRNA sequence.
ACCESSION AA928608
VERSION AA928608.1 GI:3076899
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 282)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
```

```
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 846 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 255.
Location/Qualifiers
1. .282
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1552973"
/clone_lib="NCI_CGAP_GC4"
/tissue_type="pooled germ cell tumors"
/lab_host="DHI08"
/Note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified p7T3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
```

FEATURES

```
source
BASE COUNT 63 a 68 c 66 g 84 t 1 others
ORIGIN

Query Match 0.9%; Score 228; DB 13; Length 282;
Best Local Similarity 100.0%; Pred. No. 3.3e-66;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26238 gcacagcgcggcgagggctgccacacgaaagcagagatgaactatttcattc 26297
|
|
|
Db 228 GCACAGCGCGGCGAGGCTGCCACACGAGCAAGCAGATGAATTAATTTTCATTTC 169
|
|
|
Qy 26298 aaggcagtttttaagaagcttggaaacagcggcgacatttcctaatccagcaa 26357
|
|
|
Db 168 AAGGCAGTTTTTAAGAAAGTCTTGGAAACAGCAGCGCGACCTTTCCTCTAATCCAGCAA 109
|
|
|
Qy 26358 agtgattccctgcacaccagagacaagcagagtaacagatcagtggtcctaagtgtccg 26417
|
|
|
Db 108 AGTGATTCCTTCGACACACAGAGCAGAGTACAGATCAGTGGGCTCAAGTGTCG 49
|
|
|
Qy 26418 agacttaacgaaaatagttatttcagctgcaataaagattgagttgca 26465
|
|
|
Db 48 AGACTTAACGAAAATAGTATTTCAGCTGCAATAAAGATTGAGTTTGC 1
```

```
RESULT 53
AI357786/c 433 bp mRNA EST 15-FEB-1999
LOCUS q98d07.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:1980109 3',
DEFINITION mRNA sequence.
ACCESSION AI357786
VERSION AI357786.1 GI:4109407
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 433)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
```



```

VERSION      AW247380.1  GI:5590373
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 249)
AUTHORS      NIH-MGC http://mgi.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Other_ESRs: 2820640.3prime
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF cDNA Library Preparation: Ling
Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E.
Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing
project Clone distribution: MGC clone distribution information can
be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html Base Calling / Quality
Scores: PHRED from University of Washington Genome Center. Vector
Trimming: cross_match from University of Washington Genome Center
PRAP suite. Poly-T Identification: patMatch.pl from Berkeley
Drosophila Genome Project. University of Washington Genome Center:
http://www.genome.washington.edu
Plate: LLCM4 row: L column: 17
High quality sequence stop: 225.

FEATURES             source
    source
1. 249
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:2820640"
    /clone_lib="NIH_MGC.7"
    /tissue_type="small cell carcinoma"
    /cell_line="MGC3"
    /lab_host="DH10B (phage-resistant)"
    /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
    EcoRI; cDNA made by oligo-dT priming. Directionally
    cloned into EcoRI/XhoI sites using the following 5'
    adaptor: GCACGAG(G). Size-selected >500bp for average
    insert size 1.8kb. Library constructed by Ling Hong in
    the laboratory of Gerald M. Rubin (University of
    California, Berkeley) using ZAP-cDNA synthesis kit
    (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT          39 a 83 c 94 g 33 t
ORIGIN

Query Match          0.8%; Score 221; DB 113; Length 249;
Best Local Similarity 100.0%; Pred. No. 7.7e-64;
Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 850 gcacggaaacgcggcgtagtgaccgcggctttctcagtttggtggagacggcgc 909
      |||||||
Db 10 GCACGGGAAACCGGGCGGTAGTGACCGCGCTTCTCAGTTTGTGGAGACGGGCGC 69

Qy 910 atgtggcgctttgctgctgctgcggtccgcgcgcgcacccatgtcgcaggacgc 969
      |||||||
Db 70 ATGTGGCGCTTGTGCTGCTGCTGGGTCCGGGCGCGGACGACCATGTGCGAGGACGC 129

Qy 970 accatatcgaggacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcctg 1029
      |||||||
Db 130 ACCATATCGCAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 189

Qy 1030 qcacgcagagaaagcgcggaccgtcggtgctcgcgcgcgcgcgcgcgcgcgcgc 1070
      |||||||
Db 150 CGCACGCCGAGAGACGGCGGACCGCGGCGGTGCTCCGGCGG 230

RESULT 56
AI804749/c
LOCUS      tu42d02.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2253699 3',
DEFINITION mRNA_sequence.
```

```

ACCESSION    AI804749
VERSION      AI804749.1  GI:5370221
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 477)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL      Tumor Gene Index
COMMENT      Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 876 Std Error: 0.00
Seq primer: -400p from Gibco
High quality sequence stop: 458.

FEATURES             source
    source
1. 477
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:2253699"
    /clone_lib="NCI_CGAP_Pr28"
    /sex="male"
    /dev_stage="adult"
    /lab_host="DH10B"
    /note="Organ: prostate; Vector: p7T3D-pac (Pharmacia)
    with a modified polylinker; Plasmid DNA from the
    normalized library NCI_CGAP_Pr22 was prepared, and ss
    circles were made in vitro. Following HAP purification,
    this DNA was used as tracer in a subtractive hybridization
    reaction. The driver was PCR-amplified cDNAs from a pool
    of 5,000 clones made from the same library (cloneIDs
    985608-986759, 1101192-1101959, and 1217928-1220615).
    Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT          93 a 130 c 120 g 133 t 1 others
ORIGIN

Query Match          0.8%; Score 220; DB 102; Length 477;
Best Local Similarity 98.9%; Pred. No. 1.3e-63;
Matches 470; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 25992 cgggccacacagaggagccagccagaggtcagagccagtcgaagatctggaga 26051
      |||||||
Db 477 CGGGCCACACAGAGAGCCACAGCCAGAGAGGTTCAGAGCCAGTCGAGATCTGGAGA 418

Qy 26052 ccctgaactcagaagctgtgtcttctgccccacgcacgccctatctgcctct 26111
      |||||||
Db 417 CCCTGAACTCAGAGGCTGTGTCTTCTGCCCCACGCACGCCGATCTGCCCTCT 358

Qy 26112 tgcctgtagaagctgaagagcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 26171
      |||||||
Db 357 TGCTGGTAAAGCTGAAGAGCAGCGTCCCGCCAGAGGAGCTCAGAGTAGGTGTATGGA 298

Qy 26172 gctgtgcagagcttggtctccacataagcactagtctatagctctttaggactgg 26231
      |||||||
Db 297 GCTGTGCCNAGGCTTGGGGTCCCATATAGCACTAGTCTATATAGTGCCTCTTAGGACTGG 238

Qy 26232 tgcctggcacagccgcggccaggaggtgccacacgcagcaagcaagcagatgaactaatt 26291
      |||||||
Db 237 TGCCTGGCACAGCTCGCGGCCAGGAGGTGCCACACGGAAGCAACGACAGTGAATTT 178

Qy 26292 catttcaggcaggtttttaagaaagctcttgaaacagacgcgcgcgccttctctaatc 26351
      |||||||
```

Db	256	CCCAGATTATCAATTTGGATTATTTTGAACNGAATTAATTTTCTTCCATTAACCTGCATTTG	197
Qy	14229	aaacaaatgagtggtggtcagagtggtgagactgtcgtggtcaagagtcggtgttatggg	14288
Db	196	AAACAAATGAGTGGGTGAGAGTGTGAGACTGTCGTGGTCAAGAGTCCGTGTTATGGG	137
Qy	14289	atgactcaacagctggggaatgtcttttgggctaactgcactcgtgtgtgtcctctat	14348
Db	136	ATGACTCACAGCTGGGGAATGTCTTTGTGCTAACTGACACTCTGTGTGTCTCTAT	77
Qy	14349	cgaagttaaacagtttgcggttcagctttcattccagatggaatcatctttgaccacc	14408
Db	76	CGAAGTTAAACAGTTCGCGGTTGAGCTTCATTCAGATGAATCAATCTTTGACCACC	17
Qy	14409	tatctgagttggaatc	14424
Db	16	TATCTGAGTTTGATC	1
RESULT	58		
LOCUS	AW511765/c		
DEFINITION	xu76f03.xl NCI_CGAP_Kid8 Homo sapiens EST	03-MAR-2000	
ACCESSION	AW511765		
VERSION	AW511765.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 316)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaaps-r@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emert-Buck, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. DNA sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml Seq primer: -40UP from Gibco High quality sequence stop: 239. Location/Qualifiers 1. 316 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2807645" /clone_lib="NCI_CGAP_Kid8" /tissue_type="renal cell tumor" /lab_host="DH10B" /note="Organ: kidney; Vector: pCMV-SPORT6; Site:1; SalI; Site:2; NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.2 kb. Life Technologies catalog #: 11524-014"		
BASE COUNT	69 a	84 c	72 g 90 t 1 others
ORIGIN			
Query Match	0.88;	Score 214;	DB 117; Length 316;
Best Local Similarity	99.6%;	Pred. No. 1.6e-61;	
Matches	264; Conservative	0; Mismatches	1; Indels 0; Gaps 0;
Qy	26200	agcactagtctatagatgcctcttagactggtccctggcacacgcgcggccaggagcc	26259
Db	265	AGCATTAGCTATAGATGCCCTCTTAGACTCGTGGCGACACCGCGGCCAGGAGGC	206
Qy	26260	tqccacacggagcaagcagatgaactaatcttcattccaagcagctttttaaaadaqtct	26319


```
QY 26229 tggctgctgcacagccgcccagagagctgccacacggaagcaagcagatgaactaa 26288
|||||
Db 236 TGGTGCTGGCAGCAGTGGCGCCAGGAGGCTGCCACAGCAAGCAGATGAACFAA 177
|||||
QY 26289 ttctattcaaggcagtttttaaaagaagctctggaacacagcggcgcccttccctcta 26348
|||||
Db 176 TTTTCATTTCAAGCGAGTTTTTAAAGAAGTCATGGAACACAGACGGGGCACCTTTTCCTCTA 117
|||||
QY 26349 atccagcaaaagtattccctgcacaccagagacaagcagagtaaacagatcagtggtct 26408
|||||
Db 116 ATCCAGCAAAATGATTCCTGTCACACAGACAGACAGATGAACAGGATCAGTGGGCT 57
|||||
QY 26409 aagtgccgagacttaacgaataatagttatttcagctgcaataaagattgagttgc 26464
|||||
Db 56 AAGTGCCGAGACTTAACGAAATAGTATTTCAGCTGCAATAAAGATTGAGTTTGC 1

RESULT 63
AQ349651/c
LOCUS
DEFINITION
  AQ349651      220 bp      DNA      GSS      07-MAY-1999
  RPCI11-126L18-TV RPCI-11 Homo sapiens genomic clone RPCI-11-126L18,
  DNA sequence.
ACCESSION
  AQ349651
VERSION
  AQ349651.1 GI:4176986
KEYWORDS
  GSS.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 220)
  Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
  ,J.C.
  Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
  Map Building
  Unpublished (1997)
  Other_GSSs: RPCI11-126L18.TJ
  Contact: Shaying Zhao, William Nierman, Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: hbe@tigr.org
  Clones are derived from the human BAC library RPCI-11. For BAC
  library availability, please contact Pieter de Jong
  (pieter@dejong.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
  Research Genetics (info@resgen.com). BAC end search page:
  http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
  Seq primer: T7
  Class: BAC ends.
FEATURES
  source
    1..220
    /organism="Homo sapiens"
    /db_xref="GDB:7548281"
    /db_xref="taxon:9606"
    /clone="RPCI-11-126L18"
    /sex="Male"
    /cell_type="Lymphocytes"
    /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
    RPCI11 Human Male BAC Library"
BASE COUNT
  64 a 43 c 67 g 46 t
ORIGIN

Query Match 0.8%; Score 210; DB 227; Length 220;
Best Local Similarity 100.0%; Pred. No. 4e-60;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15459 ccagggctacacacctctgcacctgcctgaactttctgtgggattcctgcctccacag 15518
|||||
Db 210 CCAGCGCTACCTTACCCTTCTGCACCTGCTAAACTTTCTGTGGGATTCCTGCCTTCCAG 151
```

```
QY 15519 aattcaggtccacagatctgtctacactcgtgaagaaatgcacgcctaggtggcgc 15578
|||||
Db 150 AATTCTAGCTTCCCAGATCTGTCTACACTCGTGAAGAAATGCACCCCTAGGTGGCCG 91
|||||
QY 15579 agtgccacacagattccatttatttacacccctccacactcttcagggtgtctgaacaaa 15638
|||||
Db 90 AGTGTCACACAGCATTCATTTATTTTACACCCCTCCACACTCTTCAGGGTGTCTGAACAAA 31
|||||
QY 15639 tactcgcctgttggtgagagattccataagt 15668
|||||
Db 30 TACTCCCGTTGGTTGAGGATTCCTAAGT 1

RESULT 64
AW175582/c
LOCUS
DEFINITION
  AW175582      225 bp      mRNA      EST      16-NOV-1999
  QVO-BT0041-030999-013-e09_1 BT0041 Homo sapiens cDNA, mRNA
  sequence.
ACCESSION
  AW175582
VERSION
  AW175582.1 GI:6441619
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 225)
  HCCP http://www.ludwig.org.br/ORESTES.
  The FAPESP/LICR Human Cancer Genome Project
  Unpublished (1999)
  Contact: Simpson A.J.G.
  Laboratory of Cancer Genetics
  Ludwig Institute for Cancer Research
  Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
  Brazil
  Tel: +55-11-2704922
  Fax: +55-11-2707001
  Email: asimpson@ludwig.org.br
  This sequence was derived from the FAPESP/LICR Human Cancer Genome
  Project. This entry can be seen in the following URL
  (http://www.ludwig.org.br/scripts/gethtml2.pl?l1-QV0&t2-QV0-BT0041-
  030999-013-e09_1&t3-1999-09-03&t4-1)
  Seq primer: puc 18 forward
  High quality sequence stop: 224.
FEATURES
  source
    1..225
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="BT0041"
    /dev_stage="Adult"
    /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
    SmaI; A mini-library was made by cloning products derived
    from ORESTES PCR (U.S. Letters Patent application No. 196
    ,716 - Ludwig Institute for Cancer Research) profiles
    into the pUC 18 vector. Reverse transcription of tissue
    mRNA and cDNA amplification were performed under low
    stringency conditions."
BASE COUNT
  78 a 66 c 41 g 40 t
ORIGIN

Query Match 0.8%; Score 209; DB 112; Length 225;
Best Local Similarity 100.0%; Pred. No. 8.7e-60;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13632 ttgtgctttgtttacctctcagctcactgagatgatcatgttagcagttgcctt 13691
|||||
Db 209 TTTGTGCTTTGTTTACCTCTCAGCTCAGCTAGGATATGTCATGTAGCAGTTCCTT 150
|||||
QY 13692 gaaggcaattcagtttggctgagctgtgacccccagtgggcggttattggtttt 13751
|||||
Db 149 GAAGGCATTCAGTTTGGTTGGCTGAGCTGTGACCCCCAGTGGCGGGCTTATTGTGTTT 90
```

```
QY 13752 gcagatttggcgaagcgtgtactcctccagatcctggctgtcttttgggtggt 13811
|||||
Db 89 GCAGATTTGGCTGAAGAGCTGTGTACTCTCCAGATCCTGGTGTCTTTTGGTGGT 30
|||||

QY 13812 agaatgccagatgaagcgttcattcaac 13840
|||||
Db 29 AGAATGCCAGATGAAGGCTTCATTCAAC 1
|||||

RESULT 65
AA504146/c 219 bp mRNA EST 18-AUG-1997
LOCUS aa59e06.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:825250 3',
DEFINITION mRNA sequence.
ACCESSION AA504146
VERSION AA504146.1 GI:2240306
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 219)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40ml3 fwd. RT from Amersham
High quality sequence stop: 175.
FEATURES
Location/Qualifiers
source
1..219
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:825250"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DHI0B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGGCGCGCTCATTTTTTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

BASE COUNT 48 a 46 c 43 g 82 t

Query Match 0.8%; Score 207; DB 8; Length 219;
Best Local Similarity 100.0%; Pred. No. 4.1e-59;
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 26260 tgcacacgaagcagcagatgaactatttcatttcacggcagtgcttttaagaagtct 26319
|||||
Db 219 TGCCACACGGAAGCAGCAGATGAACTAATTCATTTCAAGGCGAGTTTAAAGAAGTCT 160
|||||

QY 26320 tggaaacagacggcgccaccttctcctaatacceaagaagtattccctgcacaccagag 26379
|||||
```

```
Db 159 TGGAAACAGACGGCGGCACCTTCTCTTAATCCAGCAAGTGATTCCTCGCACACCAGAG 100
|||||

QY 26380 acaagcagagtaaacagatcagtggtgcttaagtctcagagacttaacgaaaaatagtattt 26439
|||||
Db 99 ACAAGCAGAGTAACAGGATCAGTGGGTCTAAGTCTCCGAGACTTAAACGAAAATAGTATTT 40
|||||

QY 26440 cagctgcaataaagattgagtttcaa 26466
|||||
Db 39 CAGCTGCAATAAAGATTGAGTTTGCAA 13
|||||

RESULT 66
BF477438/c 409 bp mRNA EST 05-DEC-2000
LOCUS nac60h05.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3438945
DEFINITION 3', mRNA sequence.
ACCESSION BF477438
VERSION BF477438.1 GI:11548265
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 409)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTCAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco.
FEATURES
Location/Qualifiers
source
1..409
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3438945"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DHI0B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGCGCGCATATCTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."
```

BASE COUNT 88 a 110 c 101 g 110 t

Query Match 0.8%; Score 205; DB 149; Length 409;
Best Local Similarity 99.0%; Pred. No. 1.6e-58;
Matches 405; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 26056 gaactcagaaggctgtgtgtcttctgtcccaacgacgcacccgtatctgaccttgcct 26115
|||||
Db 409 GAACTCAGAAGGCTGTGTGTCTTCTGCCCCACGCACCCGATCTGCCCTCTTGCCT 350
|||||

QY 26116 ggttagagcctgaagacgacggtccccccagagcgagctcaggtaggtggatgagctg 26175
|||||
```

```
Db 349 GGTAGAAGCTGAAGACGACCGCTCCCGGAGGAGCGAGCTCAGGATAGGTGGTATGGAGCTG 290
Qy 26176 tgcagagcttgggctcccaataagcactagtctatagatgcctcttagaactggtgcc 26235
Db 289 TGCCGAGCTTGGGGTCCCAATAGCACTAGTCTATAGATGCCCTCTTAGGACTGGTGCC 230
Qy 26236 tggcacagcccgggccaggaggtgccacacggaagcaagcagatgaactaatttcatt 26295
Db 229 TGGCACAGCTGCGGCCAGGAGGCTGCCACACGGAAGCAAGCAGATGAACTAATTTTCATT 170
Qy 26296 tcaagcgactttttaagaagtcttggaaacagacgcgccaccttctcctaataccagc 26355
Db 169 TCAAGGCGAGTTTTTAAAGAAGTCATGGAACAGAGCGGGCGGACCTTTCTCTTAATCCAGC 110
Qy 26356 aaagtgtacctgcacaccagagacagcagagtaaacagatcagtggtcttaagtctc 26415
Db 109 AAATGATTCCTCGCACACCAGAGACAAGCAGAGTAACAGGATCAGTGGGTCTAAGTGTG 50
Qy 26416 cgagacttaacgaaaaatagttttcagctgcaataaagattgagtttgc 26464
Db 49 CGAGACTTAACGAAAAATAGTATTTTCAGCTGCAATAAAGATTGAGTTGC 1

RESULT 67
AUI48489/c
LOCUS AUI48489 541 bp mRNA EST 25-OCT-2000
DEFINITION AUI48489 NT2RM4 Homo sapiens cDNA clone NT2RM4000375 3', mRNA
sequence.
ACCESSION AUI48489
VERSION AUI48489.1 GI:11010010
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 541)
AUTHORS Ota,T., Wakamatsu,A., Nishikawa,T., Ishii,S., Yamamoto,J., Nakamura
,Y., Nagai,T., Suzuki,Y., Sugano,S. and Isogai,T.
TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Nishikawa,T., Ishii
,S., Yamamoto,J., Nakamura,Y., Nagai,T., Suzuki,Y., Sugano,S.,
Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1. .541
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM4000375"
/clone_lib="NT2RM4"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/notes="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"
BASE COUNT 107 a 150 c 135 g 141 t 8 others
ORIGIN
Query Match 0.8%; Score 204; DB 108; Length 541;
Best Local Similarity 99.3%; Pred. No. 3.1e-58;
Matches 304; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 26013 caggccaagaaggtcagagccagtcgaagatctgggagaccctgaactcagaagctgtg 26072
, |||||
```

```
Db 452 CAGGCCAAGAAGGTGAGAGCCCGAGTGAAGATCTGGGAGACCCCTGAACCTCAGAGGCTGTG 393
Qy 26073 tgtcttctgccccacgacgaccccgctatctgcccctcttgcgtgtagaagctgaagac 26132
Db 392 TGCTCTCTGCCCCACGACGACCCGCTATCTGCCCCCTCTTGTGTTGAAGCTGAAGAGC 333
Qy 26133 acggttccccagggagcgagctcagataggttggtatggagctgtgccgaggttgggtcc 26192
Db 332 ACGTGTCCCGAGGAGGAGCTCAGGATAGGTGGTATGGAGCTGTGCCGAGGCTTGGGTGTC 273
Qy 26193 caacataagcactagtctatagatgcctcttaggactggtgctgcacagccgagggcc 26252
Db 272 CCACATAAGCACTAGTCTATAGATGCCCTCTTAGGACTGTGGTCTGGCACAGCTGCGGGCC 213
Qy 26253 aggaggtgcccacacgagcaagcaagcagatgaactaatttcatttcagcgagctttttaa 26312
Db 212 AGGAGGCTGCCACACGGAAGCAAGCAGATGAACTAATTTTCATTTCAAGGACAGCTTTTAAA 153
Qy 26313 gaagtc 26318
Db 152 GAAGTC 147

RESULT 68
AUI41263/c
LOCUS AUI41263 416 bp mRNA EST 05-OCT-1998
DEFINITION AUI41263 q46h05.sl Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1689849
3', mRNA sequence.
ACCESSION AUI41263
VERSION AUI41263.1 GI:3648720
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 416)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 711 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 233.
FEATURES
source
1. .416
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1689849"
/clone_lib="Soares_NhHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/notes="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT 87 a 108 c 99 g 122 t
ORIGIN
Query Match 0.7%; Score 198; DB 16; Length 416;
Best Local Similarity 99.0%; Pred. No. 3.5e-56;
```


[illegible]

[illegible]

```
ORIGIN
Query Match      0.7%; Score 182; DB 157; Length 424;
Best Local Similarity 100.0%; Pred. No. 8.5e-51;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26270 aagaacagatgaactaatttcatttcaaggcaggtttttaaagaagcttggaaacaga 26329
|||||
Db 195 AAGCAAGCAGATGAACATAATTTCAATTTTCAAGCGAGTTTTTAAAGAAGTCTTTGGAAACAGA 136
|||||

Qy 26330 cggcgccaccttctctaatccagcaagtgattccctgcacaccagacagacag 26389
|||||
Db 135 CGCGCGCACCTTCTCTAATCCAGCAAGTGATTCCTCGCACACAGACAGACAGAG 76
|||||

Qy 26390 taacagagatcagtggtggtcagtgccgagacttaacgaaatagttattcagctgcaat 26449
|||||
Db 75 TAACAGGATCAGTGGTCTAAGTGCCGAGACTTAACGAAATAGTATTTACGCTGCAAT 16
|||||

Qy 26450 aa 26451
||
Db 15 AA 14

RESULT 75
A1033342/c
LOCUS      A1033342      482 bp      mRNA      EST      28-AUG-1998
DEFINITION      ox02d12.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
clone IMAGE:1655159 3', mRNA sequence.
ACCESSION      A1033342
VERSION      A1033342.1 GI:3254295
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 482)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 749 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 455.
FEATURES
source
1..482
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1655159"
/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
This is a modified polylinker; Site.1: Pac I; Site.2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen 1NFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'
AATCGGAAGTAATTAATTAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Ronaldo."
94 a 130 c 122 g 136 t

BASE COUNT
ORIGIN

Query Match      0.7%; Score 180; DB 15; Length 482;
```

```
Best Local Similarity 100.0%; Pred. No. 3.8e-50;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26010 ccacagggccaagaaggtcagagccagtgaaagatctggagaccctgaactcagaaggtc 26069
|||||
Db 456 CCACAGGCCAAGAAGGTCAGAGCCAGTGAAGATCTGGAGACCCCTGAACCTCAGAAGGCT 397
|||||

Qy 26070 gtgtgtcttgcgcccccgcagcgcaccccgatctctgcctcctctgtgtagaagctgaag 26129
|||||
Db 396 GTGTGTCCTTGTCCCCACGACGACCGGTATCTGCCCTCCTTGTGGTAGAAGCTGAAG 337
|||||

Qy 26130 agcaggttccccccagagagcagctcagataggtgtgtatgagctgtaccagagcttggg 26189
|||||
Db 336 AGCAGGTGCCCCAGGAGGACCTCAGGATAGGTGTGTGGAGCTGTGTCGGAGGCTTGGG 277
|||||

RESULT 76
BE619874/c
LOCUS      BE619874      577 bp      mRNA      EST      20-OCT-2000
DEFINITION      601473130T1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3876223 3',
mRNA sequence.
ACCESSION      BE619874
VERSION      BE619874.1 GI:9890812
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 577)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
JOURNAL
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/BTP/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9636 row: k column: 08
High quality sequence stop: 577.
FEATURES
source
1..577
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3876223"
/clone_lib="NIH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SpORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
94 a 189 c 145 g 149 t

BASE COUNT
ORIGIN

Query Match      0.7%; Score 179; DB 138; Length 577;
Best Local Similarity 100.0%; Pred. No. 7.8e-50;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26102 ctgcctcttctggttagagctgaagcacggtccccccagagggcagctcagagtag 26161
|||||
Db 323 CTGCCCTCTTCTGTGTAGAGCTGAAGACACGGTCCCCCAGGAGGAGCTCAGAGTAG 264
|||||

Qy 26162 gtgtatggagctgtgccgaggttgggtctccacataagcactagtctatagatgcctc 26221
|||||
Db 263 GTGTATGGAGCTGTGCCGAGGCTTGGGCTCCACATAAGCAC TAGTCTATAGATGCCTC 204
|||||

Qy 26222 ttaggactgtgtccttgccacagccgaggaggtgccacacggaagcagaga 26280
|||||
```



```
|||||
Db 106 TAACAGGATCAGTGGGTCTAGAGTCCGAGACTTAACGAAATAGTATTTCAGCTGCAAT 47
|||||
Qy 26450 aaagattgattgcaa 26466
|||||
Db 46 AAGATTGAGTTGCAA 30
|||||

RESULT 79
AA905284/c 222 bp mRNA EST 19-MAY-1998
LOCUS OJ96b07.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:I506133.3', mRNA sequence.
ACCESSION AA905284
VERSION AA905284.1 GI:3040407
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 222)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 865 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 213.

FEATURES
source
1. 222
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:I506133"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NDHL19W, testis NHT, and B-cell
NCI-CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 47 a 54 c 48 g 73 t
ORIGIN

Query Match 0.6%; Score 171; DB 13; Length 222;
Best Local Similarity 99.5%; Pred. No. 5.3e-47;
Matches 221; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 26245 cgcgggcaggaggtgccacacggaagcaagcagatgaacttaatttcattcaggcag 26304
|||||
Db 222 CGCGGGCCAGGAGGTGCCACACGGAAGCAAGCAGATGAACCTAATTTTCATTCAAGGCAG 163
|||||

Qy 26305 tttttaagaagttcttggaaacagacgagcggcgccaccttcctccttaataccagcaaatgatt 26364
|||||
Db 162 TTTTAAAGAAGTCATGTAAGAACAGACGCGGCACCTTTTCCTTAATCAACAGAAAGTGATT 103
|||||

Qy 26365 ccctgcaccacagacagacagatgaatgaatcagtggtgtcgaagtcgcgagactta 26424
|||||
Db 102 CCCTGCACACACAGACAGCAGAGTACAGATCAGTGGGTCTAAGTGTCCGAGACTTA 43
|||||

Qy 26425 acgaaatagttatttcagctcgaataaagattgagttgcaa 26466
|||||
```

```
Db 42 ACGAAATAGTATTTCAGCTGCAATAAAGATTGAGTTTGCAA 1
|||||
RESULT 80
AA766184/c 318 bp mRNA EST 08-FEB-1998
LOCUS oal2f09.s1 NCI-CGAP_GCB1 Homo sapiens cDNA clone IMAGE:I304777.3',
DEFINITION mRNA sequence.
ACCESSION AA766184
VERSION AA766184.1 GI:2817422
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 318)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1277 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 288.

FEATURES
source
1. 318
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:I304777"
/clone_lib="NCI-CGAP_GCB1"
/lab_host="DH10B"
/tissue_type="germinal center B cell"
/notes="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTAGCAATCTGAAGTGGGAGGCGCGCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 71 a 82 c 70 g 95 t
ORIGIN

Query Match 0.6%; Score 171; DB 11; Length 318;
Best Local Similarity 100.0%; Pred. No. 4.7e-47;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26156 ggatagtggtatggagctgtgcagagcttgggtccacataagaactagtctataga 26215
|||||
Db 318 GGATAGTGGTATGGAGCTGTGCCAGGCTTGGGCTCCACATAAGCTAGTCTATAGA 259
|||||

Qy 26216 tgcctcttagactggttcctggcaagcagccgcccaggaggtgccacacggaagcaa 26275
|||||
Db 258 TGCCTCTTAGACTGCTGCTGGCACACCGCGGCGCAGAGGCTGCCACAGGAGCAA 199
|||||

Qy 26276 gcagatgaactaatttcatttcaggcgagtttttaagaagtccttggaaac 26326
|||||
```


RESULT 83
AI500718/c 372 bp mRNA EST 14-APR-1999
LOCUS tn94b10.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2177179 3',
DEFINITION mRNA sequence.
ACCESSION AI500718
VERSION AI500718.1 GI:4392700
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 372)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CNA Library Preparation: Life Technologies, Inc.
CNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 737 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 342.
FEATURES
source
1..372
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2177179"
/clone_lib="NCI_CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.85 kb. Life Technologies catalog #: 11539-012"
BASE COUNT 80 a 100 c 82 g 110 t
ORIGIN
Query Match 0.6%; Score 162; DB 21; Length 372;
Best Local Similarity 98.9%; Pred. No. 4.7e-44;
Matches 362; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 26107 ctcttgctgttagaagctgaagacacggtccccccagaggagcagctcagatagtggt 26166
Db 372 CTCCTTGCTGTGTAAGCTGAAGACGACGGTCCCCCAGGAGCGACTCAGGATAGGTGGT 313
Qy 26167 atggagctgtgccagggcttgggtccccacataagcactagtctatagatgctcttagg 26226
Db 312 ATGGAGCTGTGCCAGGGCTTGGGTCCCCACATAGCCTAGTCTATAGATGCCCTTAGG 253
Qy 26227 actggtgcttgccacagccgcgggccaggaggtgctgccacacggaagcagatgaact 26286
Db 252 ACTGTTGCTGTCACAGCTGCGGGCCAGGAGGTGCCACACGGAAGCAGATGAAC 193
Qy 26287 aatttcattcaaggcagtttttaagaagctgtggaacacagcgcgcaccccttcctc 26346
Db 192 AATTTCATTCAAGGCAGTTTTTAAGAAGTCATGGAACACAGCGCGCACCTTTCCTC 133
Qy 26347 taatccagcaagtgtattccctgcacacacagacagaacagcagatgaacaggtcagtggt 26406
Db 132 TAATCCACCAAAATGATTCCTCGCACACCAGACAGACAGATGAACAGGATCAGTGGGT 73
Qy 26407 ctaagtgtccagacttaacgaaaaatagatttcagctgcataaagattgagtttgcaa 26466
|||||

Db 72 CTAAGTGTCCGAGACTTAACGAAATAAGTATTTCAGCTGCAATAAAGATTGATTGCAA 13
Qy 26467 ttgtga 26472
Db 12 TTGTGA 7
RESULT 84
AA223338/c 374 bp mRNA EST 12-MAR-1998
LOCUS AA223338
DEFINITION Zr05h05.sl Stratagene NT2 neuronal precursor 937230 Homo sapiens
cDNA clone IMAGE:650649 3', mRNA sequence.
ACCESSION AA223338
VERSION AA223338.1 GI:1843862
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 374)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 3011 Std Error: 0.00
Seq primer: -4lm13 fwd. ET from Amersham.
FEATURES
source
1..374
/organism="Homo sapiens"
/db_xref="GDB:5277238"
/db_xref="taxon:9606"
/clone="IMAGE:650649"
/clone_lib="Stratagene NT2 neuronal precursor 937230"
/tissue_type="neuroepithelial cells"
/dev_stage="Ntera-2 neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: brain; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Uninduced, exponentially growing neuroepithelial
cells (Ntera-2/Cl.D1). Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."
BASE COUNT 78 a 98 c 85 g 112 t 1 others
ORIGIN
Query Match 0.6%; Score 161; DB 4; Length 374;
Best Local Similarity 98.9%; Pred. No. 1e-43;
Matches 361; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 26102 ctgcccctcttgctgtgtagaagctgaagacacggtccccccagggagcagctcagatag 26161
Db 374 CTGCCCTCTTGTGTAGAGCTGAAGACGACGGTCCCCCAGGAGGAGCTCAGGATAG 315
Qy 26162 gtggtatgagctgtgccgaggttgggtccccacataagcactagtctatagatgcctc 26221
Db 314 GTGGTATGGAGCTGTGCCAGGCTTGGGNTCCACATAAGCACTAGTCTATAGATGCCTC 255
Qy 26222 ttgagctgtgctctggcacacgcgcggcagagctgccacacgaagcagatgaact 26281
Db 254 TTAGAGCTGGTGGCTGGCCAGCAGCTGCGGGCCAGGAGGTGCCACACGGAAGCAGAT 195
Qy 26282 gaactaatttcattcaaggcaggtttttaagaagctcttggaacacagcgcgcacacct 26341
|||||

```

Db 194 GAACTAATTTTCATTTCAAGCAGTTTTTAAAGAGTCAATGAAACACGCGGCACCTT 135
|||||
QY 26342 tcctctaatccagcaaaagtatccctgcacaccagagacaagcagaggttaacaggatcag 26401
|||||
Db 134 TCCTCTAATCCAGCAAAATGATTCCTTCACACACAGAGAGAGAGAGTAAAGATCAG 75
|||||
QY 26402 tgggtctaaagtctccgagacttaacgaaatagttatttcagctgcaataaagattgatt 26461
|||||
Db 74 TGGGTCTAAGTGTCCGAGACTTAACGAAATAGTATTTCAGCTGCAATAAAGATTGATT 15
|||||
QY 26462 tgcga 26466
|||||
Db 14 TGCAA 10
|||||

RESULT 85
AI468143 421 bp mRNA EST 30-MAR-1999
LOCUS
DEFINITION
tf92905.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2106776 3'
similar to SW:YK59_YEAST P36159 HYPOTHETICAL 96.8 KD PROTEIN IN
SIS2-MTD1 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION
AI468143 GI:4330233
VERSION
AI468143.1
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 421)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/NTGAP), Tumor Gene Index
Unpublished (1998)
JOURNAL
Contact: Robert Strausberg, Ph.D.
COMMENT
Email: cgapb-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert length: 733 Std Error: 0.00
Seq primer: -400P from Gibco.
FEATURES
Location/Qualifiers
source
1..421
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2106776"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGCATATCTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 98 a 114 c 121 g 88 t
ORIGIN

Query Match 0.6%; Score 158; DB 20; Length 421;
Best Local Similarity 100.0%; Pred. No. 1e-42;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 999 gcgccgcgcaagaccgcctgcggcacctgcgcacgcagagagcgccgacccgtcg 1058
|||||
Db 1 GCGGCCGCGCAAGACCCGCTGCGGACCTGCGCACGAGAGAGCGCGACCGTCGGG 60
|||||
QY 1059 gtgctcggcgcccaaacaccgtgtacctgcaggttggtgcagcggttagccgggactc 1118
|||||
Db 61 GTGCTCCGGCGGCCCAACACCGGTGTACCTGTCAGTGTGGCAGCGGTAGCCGGGACTC 120
|||||
QY 1119 gggcgccgcctctacgtctctctcccgaggttcaaccggt 1156
|||||
Db 121 GGGCGCCGCGCTCTACGTCTTCTCCGAGTTCACCGGT 158
|||||

RESULT 86
AI200296 451 bp mRNA EST 14-OCT-1998
LOCUS
DEFINITION
qf86b12.x1 Soares_fetal_lung_NbHL19w Homo sapiens cDNA clone
IMAGE:1756895 3' similar to SW:YK59_YEAST P36159 HYPOTHETICAL 96.8
KD PROTEIN IN SIS2-MTD1 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION
AI200296 GI:3752902
VERSION
AI200296.1
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 451)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
Contact: Robert Strausberg, Ph.D.
COMMENT
Email: cgapb-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Gibco
High quality sequence stop: 442.
FEATURES
Location/Qualifiers
source
1..451
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1756895"
/clone_lib="Soares_fetal_lung_NbHL19w"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTTTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library. Soares
fetal heart NbHL19w."
BASE COUNT 110 a 120 c 128 g 93 t
ORIGIN

Query Match 0.6%; Score 158; DB 17; Length 451;
Best Local Similarity 100.0%; Pred. No. 9.8e-43;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 999 gcgccgcgcaagaccgcctgcggcacctgcgcacgcagagagcgccgacccgtcg 1058
|||||
Db 1 GCGGCCGCGCAAGACCCGCTGCGGACCTGCGCACGAGAGAGCGCGACCGTCGGG 60
|||||
QY 1059 gtgctcggcgcccaaacaccgtgtacctgcaggttggtgcagcggttagccgggactc 1118
|||||
Db 61 GTGCTCCGGCGGCCCAACACCGGTGTACCTGTCAGTGTGGCAGCGGTAGCCGGGACTC 120
|||||

```


Db 121 GCGCGCGGCTCTAGCTTCTTCCGAGTTCAACCGGT 157

RESULT 91
BF434169

LOCUS BF434169 553 bp mRNA EST 29-NOV-2000
DEFINITION 7099f04.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3644670 3' similar to TR:Q9V5J4 Q9V5J4 G3298 PROTEIN. ;, mRNA sequence.
ACCESSION BF434169
VERSION BF434169.1 GI:11446441
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 553)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 507.
FEATURES
source
1..553
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3644670"
/clone_lib="NCI_CGAP_Ov18"
/tissue_type="fibrothecoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCGACATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaudo. "BASE COUNT 143 a 150 c 153 g 107 t
ORIGIN

Query Match 0.6%; Score 157; DB 148; Length 553;
Best Local Similarity 100.0%; Pred. No. 2e-42;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1000 cggcgccgaagaccgctcggcactcgcacgcgacgagagaagcgcgaccctcgagg 1059
|||||
Db 1 CGGCGCGCAAGGACCGCTCGCGACCTGCGCACGCGAGAGAGCGCGGACCGTCGGGG 60
|||||

Qy 1060 tgcctccgcgccgcaaaacaccgtgtacctgcaggtgggtggcagcggttagcgggactcg 1119
|||||

Db 61 TGCTCCGCGGCCCAACACCGTGTACTCGAGGTGGTGGCAGCGGTAGCCGGGACTCG 120
|||||

Qy 1120 ggcgcgcgcgtctacgtcttctccagttcaaccggt 1156
|||||

Db 121 GCGCGCGGCTCTAGCTTCTTCCGAGTTCAACCGGT 157

RESULT 92
BF872931/c

LOCUS BF872931 354 bp mRNA EST 17-JAN-2001

CMO-ET0124-021100-676-e02 ET0124 Homo sapiens cDNA, mRNA sequence.
BF872931
VERSION BF872931.1 GI:12263061
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 354)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM0&t2=CM0&t3=CM0&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 5
High quality sequence stop: 354.
FEATURES
source
1..354
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0124"
/dev_stage="Adult"
/note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 75 a 91 c 108 g 80 t
ORIGIN

Query Match 0.6%; Score 156; DB 170; Length 354;
Best Local Similarity 100.0%; Pred. No. 5e-42;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21343 gctttgaaccagctgaagctaatgggaagtgctattgccaccttgcctccgctccgcg 21402
|||||

Db 354 GCTTTTGACACAGCTGAAGCTAATGGGAGGTGCTATTGGCCACCTTGCTCGGCTCCCG 295
|||||

Qy 21403 actcctttttccccagagaagtgatgtcttagcaccggggtctctctgcaaaatgggt 21462
|||||

Db 294 ACTCCTTTTCCCCAGAAAGGTAATGTCTTAGCACGGGGCTTCTCTCTGCAAAATGGGT 235
|||||

Qy 21463 gcagccctctcaggttcttggtctctcccccagagaa 21498
|||||

Db 234 CGAGCCCTCTCAGTGTTCGTGGCTCCTCCAGAGAA 199
|||||

RESULT 93
BF102508

LOCUS BF102508 726 bp mRNA EST 19-OCT-2000
DEFINITION 60164648F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:3931259 5', mRNA sequence.

ACCESSION BF102508
VERSION BF102508.1 GI:10885034
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL 1 (bases 1 to 726)
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW766 row: p column: 12
High quality sequence stop: 611.
Location/Qualifiers
1. 726
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3931259"
/clone_lib="NIH_MGC_60"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfII (ggccgctggcc); Site_2: SfII (ggccattatggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGCGGCGGACATG-dT(30)BN-3'
(where B = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC
Library."
BASE COUNT 170 a 198 c 219 g 139 t
ORIGIN
Query Match 0.6%; Score 155; DB 144; Length 726;
Best Local Similarity 99.6%; Pred. No. 8.5e-42;
Matches 275; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 851 cacgggaacacgagcgttagtgacgcggctttctcaattttgtggagcggcgca 910
|||||
Db 1 CACGGGAACCGGGCGGTAGTGACCGGGCTTCTCAGTTTGGTGAGACGGCGCA 60
|||||
QY 911 tgtggcgctttgctcgtcgtcggtccgcggcgccgacccatgtcgcagggacgca 970
|||||
Db 61 TGTGGCGCTTGTCTGCTGCTGGGTCCGGCGCGGACGACCATGTGCGAGGACGCA 120
|||||
QY 971 ccatacgcagggcaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1030
|||||
Db 121 CCATATCGCAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 179
|||||
QY 1031 qcacgcgagagagcggcggaccgtcggggtgctccgcggcgcccaaacacgcgtgtacatgc 1090
|||||
Db 180 GCACGCAGAGAGCGGGACCGGCGGTGCTCCGGCGGCGCCCAACACCGGTGACCTGC 239
|||||
QY 1091 aggtgggtggcagcgggtgagcgggactcggcgcccg 1126
|||||
Db 240 AGGTGGTGGCAGCGGGTAGCGGGGACTCGGGCGCGCG 275
|||||
RESULT 94
BE883616

LOCUS BE883616 370 bp mRNA EST 20-OCT-2000
DEFINITION 601508091F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3909527 5',
mRNA sequence.
ACCESSION BE883616
VERSION BE883616.1 GI:10332392
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL 1 (bases 1 to 370)
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9723 row: f column: 24
High quality sequence stop: 370.
Location/Qualifiers
1. 370
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3909527"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT 94 a 100 c 106 g 70 t
ORIGIN
Query Match 0.6%; Score 153; DB 141; Length 370;
Best Local Similarity 100.0%; Pred. No. 5.1e-41;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 25023 cagcaaacgtcccaagccatcagcgtgggatcggtatgcagcggagtcattatgct 25082
|||||
Db 125 CAGCACAAACGTCCTCAAGCCATCAGCGTGGGATGCGGATGAACGCGAGTTTCATTATGCT 184
|||||
QY 25083 gaaccacttcagccagcgcgtatgcgaaggtccctctctcagcccaacttcagcgagaa 25142
|||||
Db 185 GAACCACTTCAGCCAGCGCTATGCCAAGGTCCCTCTTTCAGCCCCCACTTCAGCGAGAA 244
|||||
QY 25143 agtgggagttgctttgaccacatgaaggtctg 25175
|||||
Db 245 AGTGGGAGTTGGCTTTGACCCACATGAAGGTCTG 277
|||||
RESULT 95
BE742908
LOCUS BE742908 677 bp mRNA EST 15-SEP-2000
DEFINITION 601574609F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3835658 5',
mRNA sequence.
ACCESSION BE742908
VERSION BE742908.1 GI:10156900
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL 1 (bases 1 to 677)
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

```
COMMENT      Contact: Robert Strausberg, Ph.D.  
             Email: cgapbs-r@mail.nih.gov  
             Tissue Procurement: DCTD/DTF  
             cDNA Library Preparation: Ling Hong/Rubin Laboratory  
             DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
             DNA Sequencing by: Incyte Genomics, Inc.  
             Clone distribution: MGC clone distribution information can be  
             found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
             Plate: LLCM518 row: a column: 03  
             High quality sequence stop: 672.  
             Location/Qualifiers  
FEATURES  
  source      1. .677  
              /organism="Homo sapiens"  
              /db_xref="taxon:9606"  
              /clone="IMAGE:3835658"  
              /clone_lib="NIH_MGC_9"  
              /tissue_type="adenocarcinoma cell line"  
              /lab_host="DH10B (phage-resistant)"  
              /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:  
              cDNA made by oligo-dT priming. Directionally  
              cloned into EcoRI/XhoI sites using the following 5'  
              adaptor: GGCACGAG(G). Size-selected >500bp for average  
              insert size 1.8kb. Library constructed by Ling Hong in  
              the laboratory of Gerald M. Rubin (University of  
              California, Berkeley) using ZAP-cDNA synthesis kit  
              (Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT   160 a 173 c 217 g 127 t  
ORIGIN  
  
Query Match      0.6%; Score 153; DB 139; Length 677;  
Best Local Similarity 100.0%; Pred. No. 4.1e-41;  
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 25023 cagcaacgtcccaagccatcagcgtgggatgagcgatgacggaggttcattatgct 25082  
Db      299 CAGCAACAACGTCCTTGACCATCAGCGTGGGATGACGGGAGTTCATTATGCT 358  
  
QY 25083 gaaccactcagcagcgtatgcgaaggtccctcttcagcccaacttcagcgagaa 25142  
Db      359 GAACCACTTCAGCCAGCGCTATGCCAAGTCCCTCTTCAGCCCACTTCAGCGAGAA 418  
  
QY 25143 agtggagtgcttctgaccacatgaaggtctg 25175  
Db      419 AGTGGGAGTGTGCTTTGACCATCATGAAGGTCTG 451  
  
RESULT 96  
BG324135  
LOCUS      BG324135      754 bp      mRNA      EST      27-FEB-2001  
DEFINITION 602423086F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4561292 5',  
            mRNA sequence.  
ACCESSION  BG324135  
VERSION     BG324135.1 GI:13130572  
KEYWORDS    EST.  
SOURCE      human.  
ORGANISM    Homo sapiens  
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE   1 (bases 1 to 754)  
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.  
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL      Unpublished (1999)  
COMMENT      Contact: Robert Strausberg, Ph.D.  
            Email: cgapbs-r@mail.nih.gov  
            Tissue Procurement: DCTD/DTF  
            cDNA Library Preparation: Ling Hong/Rubin Laboratory  
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
            DNA Sequencing by: Incyte Genomics, Inc.  
            Clone distribution: MGC clone distribution information can be  
            found through the I.M.A.G.E. Consortium/LLNL at:  
            http://image.llnl.gov  
            Plate: LLCM1271 row: k column: 21  
  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM518 row: a column: 03  
High quality sequence stop: 672.  
Location/Qualifiers  
FEATURES  
  source      1. .677  
              /organism="Homo sapiens"  
              /db_xref="taxon:9606"  
              /clone="IMAGE:3835658"  
              /clone_lib="NIH_MGC_9"  
              /tissue_type="adenocarcinoma cell line"  
              /lab_host="DH10B (phage-resistant)"  
              /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:  
              cDNA made by oligo-dT priming. Directionally  
              cloned into EcoRI/XhoI sites using the following 5'  
              adaptor: GGCACGAG(G). Size-selected >500bp for average  
              insert size 1.8kb. Library constructed by Ling Hong in  
              the laboratory of Gerald M. Rubin (University of  
              California, Berkeley) using ZAP-cDNA synthesis kit  
              (Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT   160 a 173 c 217 g 127 t  
ORIGIN  
  
Query Match      0.6%; Score 152; DB 152; Length 754;  
Best Local Similarity 100.0%; Pred. No. 8.6e-41;  
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 921 ttgtctgctgctgcggtccgcgagcgagccaccatgtcgcagggacgaccatcgca 980  
Db      70 TTGCTCGCTGCTGCGTCCGGCGCGGAGCCACCATGTGCGAGGACGCGCATATCGCA 129  
  
QY 981 ggcacccgcgcgagcgagcgccgcaagagaccgcgtgcggcaccctgcgcacgcgaga 1040  
Db      130 GGCACCCGCGCGCGCGAGCGCGCGCAAGAGACCGCTGCGGACCTGCGCAGCGAGA 189  
  
QY 1041 gaagcgagacgctcggtggtgctccgcggcc 1072  
Db      190 GAAGCGGCGACCGTCGCGGTGCTCGCGCGGCC 221  
  
RESULT 97  
BG336190  
LOCUS      BG336190      906 bp      mRNA      EST      27-FEB-2001  
DEFINITION 602404980F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4542594 5',  
            mRNA sequence.  
ACCESSION  BG336190  
VERSION     BG336190.1 GI:13142628  
KEYWORDS    EST.  
SOURCE      human.  
ORGANISM    Homo sapiens  
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE   1 (bases 1 to 906)  
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.  
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL      Unpublished (1999)  
COMMENT      Contact: Robert Strausberg, Ph.D.  
            Email: cgapbs-r@mail.nih.gov  
            Tissue Procurement: ATCC  
            cDNA Library Preparation: Ling Hong/Rubin Laboratory  
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
            DNA Sequencing by: Incyte Genomics, Inc.  
            Clone distribution: MGC clone distribution information can be  
            found through the I.M.A.G.E. Consortium/LLNL at:  
            http://image.llnl.gov  
            Plate: LLCM1222 row: p column: 19  
            High quality sequence stop: 610.  
            Location/Qualifiers  
FEATURES  
  source      1. .906  
              /organism="Homo sapiens"  
              /db_xref="taxon:9606"  
              /clone="IMAGE:4542594"  
              /clone_lib="NIH_MGC_21"  
              /tissue_type="choriocarcinoma"  
              /lab_host="DH10B (phage-resistant)"
```

/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected by average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 230 a 230 c 302 g 144 t
ORIGIN

Query Match 0.6%; Score 152; DB 152; Length 906;
Best Local Similarity 100.0%; Pred. No. 8.1e-41;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25024 agcacacgtcccaagccatcagcgtgggagtgagcgagtgagcggaggtcattatctg 25083
|||||
Db 238 AGCACACGTCCTCAAGCATCAGCGTGGGATGCGGATGAACGCGGAGTTCATTATGCTG 297
|||||

QY 25084 aaccacttcagcagcgtatgcgaagtcctccctcttcagccccaacttcagcagaaa 25143
|||||
Db 298 AACCACTTCAGCAGCGCTATGCGAGTCCCTCTTCAGCCCACTTCAGCGAGAAA 357
|||||

QY 25144 gtggaggtgcctttgaccacatgaaggtctg 25175
|||||
Db 358 GTGGAGTTCCTTTGACCATGAAGTCTG 389
|||||

RESULT 98
BF960051/c
LOCUS 157 bp mRNA EST 22-JAN-2001
DEFINITION QV2-NN0045-051200-517-h06 NN0045 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF960051
VERSION BF960051.1 GI:12377326
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV2&t2=QV2-NN0045-051200-517-h06&t3=2000-12-05&t4=1)
Seq primer: puc18 forward
High quality sequence stop: 157.
Location/Qualifiers
1. .157
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN0045"
/dev_stage="Adult"
/note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products

FEATURES
source

derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 54 a 37 c 25 g 41 t
ORIGIN

Query Match 0.5%; Score 144; DB 171; Length 157;
Best Local Similarity 100.0%; Pred. No. 7.3e-38;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6080 cagggccagatgatgtcttagcttagaagaaagtagtctgtctggaactgac 6139
|||||
Db 148 CAGGCCAGATGATGTGCTTAGCTTAGGAAAAGATTAGCTTGCTTGCTTGACCTTGCC 89
|||||

QY 6140 taaagacattcatctctgtgttttacttacatgtgaagagagtagcagaagcagtaggggtat 6199
|||||
Db 88 TAAAGACATTTCATCTCTGTTTACTTACATGTGAAGAGAGTACCAAGCAGTAGGGGTAT 29
|||||

QY 6200 ttctgttagtactaactaatgt 6223
|||||
Db 28 TTCCTGTTAGTACTAACTAATGT 5
|||||

RESULT 99
AA605004/c
LOCUS 390 bp mRNA EST 29-SEP-1997
DEFINITION no47905.s1 NCI_CGAP_Pr23 Homo sapiens cDNA clone IMAGE:1103864 3', mRNA sequence.
ACCESSION AA605004
VERSION AA605004.1 GI:2444823
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 390)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-@email.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 378.
Location/Qualifiers
1. .390
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1103864"
/clone_lib="NCI_CGAP_Pr23"
/sex="male, pooled"
/tissue_type="prostate tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: prostate; Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pooled prostate tumors. 5' adaptor sequence: 5' GAATTCGGCAGCAG 3' 3' adaptor sequence: 5' CTCGATGTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

FEATURES
source

BASE COUNT 72 a 94 c 78 g 146 t
ORIGIN

Query Match 0.5%; Score 142; DB 9; Length 390;
Best Local Similarity 98.8%; Pred. No. 2.5e-37;
Matches 342; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 26121 aagctgaagacagcagtgctcccccagagcagctcaggtatggtatgagctgtgcg 26180
|||||
Db 390 AAGCTGAAGACACGGTCCCCCAGGAGCAGCTCAGGATAGTGTGATGGAGCTGTGCCG 331
|||||

Qy 26181 aggtctggctccacataaagcactagtctatagatgctcttaggactgtgctcgca 26240
|||||
Db 330 AGGCTGGGGTCCCATATAGCACTAGTCTATAGATGCTCTTAGGACTGGTGGCGCA 271
|||||

Qy 26241 cagcgcggccagagagtgctccacacggaagcagcagatgaactaatttcattccaag 26300
|||||
Db 270 CAGCTGCGGGCAGGAGGCTGCCACACGGAAGCAGAGATGAACATAATTCATTCAAG 211
|||||

Qy 26301 gcagttttaagaagtgcttggaaacacagcggcgacaccttcctctaatccagcaagt 26360
|||||
Db 210 GCAGTTTTTAAAGAAGTCAATGAAACACAGCGCGCACCTTCTCTTAATCCAGCAAAAT 151
|||||

Qy 26361 gattccctgcacacagagacaagcagagtaacagagatcagtggtctaaagtgtccgaga 26420
|||||
Db 150 GATTCCTGTCACACAGAGACAGCAGAGTACAGATCAGTGGTCTAAAGTGCCGAGA 91
|||||

Qy 26421 cttaacgaaaatagttatttcagtgctgcaataaagattgagtttgcaa 26466
|||||
Db 90 CTTAAGCAAAATAGTATTTCAGCTGCAATAAAGATTGAGTTTGCAA 45
|||||

RESULT 100
AW806551 278 bp mRNA EST 17-MAY-2000
LOCUS ILO-ST0002-160599-003 ST0002 Homo sapiens cDNA, mRNA sequence.
DEFINITION AW806551
ACCESSION AW806551
VERSION AW806551.1 GI:7899550
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 278)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-ILO-ST0002-160
599-003&t3=1999-05-16&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 278.
Location/Qualifiers
1. .278
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ST0002"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions.*
64 a 84 c 70 g 60 t

BASE COUNT
ORIGIN

Query Match 0.5%; Score 141; DB 120; Length 278;
Best Local Similarity 100.0%; Pred. No. 6.1e-37;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16276 aggtttggcctgcacaccagcacttgctcctgaatgagaactgtgcctcagttcacacaac 16335
|||||
Db 8 AGGTTGGGGCTGCACACCCAGCAGCTTGGTCTGTAATGAGAAGTGCCTCAGTTCCACAC 67
|||||

Qy 16336 ctgcgagcccaagattcaaacccagctcaacctcatccaccggagacatcttccccctg 16395
|||||
Db 68 CTTGCGAGCCACAAGATTCAAAACCCAGCTCAACCTCATCCACCCGACATCTTCCCCCTG 127
|||||

Qy 16396 ctacccagtttcgcgtgtaag 16416
|||||
Db 128 CTCACCAAGTTTCCGCTGAAG 148
|||||

Search completed: November 2, 2001, 06:15:13
Job time: 34763 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 2, 2001, 00:26:20 ; Search time 1390.09 Seconds
(without alignments)
12044.090 Million cell updates/sec

Title: US-09-434-382-28

Perfect score: 26664

Sequence: 1 tatcagtgactgaattcta.....ttcgccaagtctttttgaca 26664

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 8

Total number of hits satisfying chosen parameters: 1388819

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : N_Geneseq_0601.*

1:	/SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.*
2:	/SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.*
3:	/SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.*
4:	/SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.*
5:	/SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.*
6:	/SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.*
7:	/SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.*
8:	/SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.*
9:	/SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.*
10:	/SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.*
11:	/SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.*
12:	/SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.*
13:	/SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.*
14:	/SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.*
15:	/SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.*
16:	/SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.*
17:	/SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.*
18:	/SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.*
19:	/SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.*
20:	/SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.*
21:	/SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.*
22:	/SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26664	100.0	26664	21	AA60207 Human prostate can
2	657	2.5	2958	21	AA58453 Human prostate can
3	514	1.9	2546	21	AA676445 Human OREF ORF2000
4	247	0.9	2478	21	AA52810 Human sulphatase G
5	158	0.6	386	22	AA64286 Novel human polynu
6	143	0.5	183	21	AA627273 Human secreted pro
7	84	0.3	50000	21	AA96365 Polymorphic repeat
8	76	0.3	17131	21	AA60888 DNA encoding a hum
9	72	0.3	238	21	AA280231 Human colon cancer
10	71	0.3	215	21	AA602520 Human secreted pro
11	60	0.2	2923	21	AA294762 Human ATP binding

c 12	60	0.2	2923	21	AA294784 Human ATP binding
c 13	60	0.2	9721	20	AA275924 Human interleukin
c 14	60	0.2	9721	21	AAF20945 Human interleukin-
c 15	60	0.2	9721	21	AA63768 Human IL-1B gene.
c 16	60	0.2	9721	21	AA50174 Human interleukin-
c 17	60	0.2	9721	21	AA50175 Human interleukin-
c 18	60	0.2	9721	21	AA34823 Human adenosine re
c 19	60	0.2	9721	22	AAF27666 IL-1B DNA. Uniden
c 20	60	0.2	9721	22	AA691434 Human IL-1B nucleo
c 21	60	0.2	10620	20	AA602996 Human IL-1ra BAC c
c 22	60	0.2	14690	20	AA22303 Human IL-1ra BAC c
c 23	60	0.2	29433	21	AA20950 Human interleukin-
c 24	60	0.2	29433	21	AA34828 Human adenosine re
c 25	60	0.2	209273	21	AAF21437 Human factor-relat
c 26	59	0.2	310	21	AA621264 Human secreted pro
c 27	59	0.2	5581	18	AA48737 Human leucocyte sp
c 28	59	0.2	5581	18	AA45451 Human LST-1 (leuko
c 29	59	0.2	21721	20	AA83427 Human lipolysis st
c 30	59	0.2	22976	20	AA83426 Genomic region con
c 31	59	0.2	23187	21	AA50273 Human lipolysis st
c 32	59	0.2	23187	22	AA62331 Human leptin fragm
c 33	59	0.2	53526	19	AA64101 Human PKD1 gene.
c 34	59	0.2	53577	17	AA18551 Human polycystic k
c 35	59	0.2	53577	19	AA64108 Human PKD1 locus b
c 36	58	0.2	2191	21	AA298085 Human secreted pro
c 37	58	0.2	2688	18	AA72165 Alzheimer's diseas
c 38	58	0.2	20674	21	AA58017 Arachidonic acid m
c 39	58	0.2	160552	22	AA602697 Human glycosyl sul
c 40	57	0.2	740	21	AA69168 Human colon cancer
c 41	57	0.2	788	21	AA61856 Human colon cancer
c 42	57	0.2	792	21	AA72064 Single nucleotide
c 43	57	0.2	792	21	AA672079 Single nucleotide
c 44	57	0.2	792	21	AA672094 Single nucleotide
c 45	57	0.2	792	21	AA672106 Single nucleotide
c 46	57	0.2	792	21	AA672109 Single nucleotide
c 47	57	0.2	792	21	AA672112 Single nucleotide
c 48	57	0.2	792	21	AA672115 Single nucleotide
c 49	57	0.2	801	21	AA61855 Human colon cancer
c 50	57	0.2	1466	20	AA225133 Human hypoxia indu
c 51	57	0.2	1559	16	AA602714 MART-1 melanoma an
c 52	57	0.2	1559	20	AA602797 Human melanoma ant
c 53	57	0.2	1559	20	AA620065 Human MART1 melano
c 54	57	0.2	183999	22	AA692831 Human ABC1 genomic
c 55	56	0.2	314	22	AA65529 Novel human polynu
c 56	55	0.2	11967	22	AA697863 Human neuroblastom
c 57	55	0.2	20303	18	AA671699 Human deoxycytidyl
c 58	55	0.2	26764	18	AA671696 Human deoxycytidyl
c 59	54	0.2	245	21	AA65670 Human differential
c 60	54	0.2	354	21	AA600358 Human secreted pro
c 61	54	0.2	723	22	AA654060 hFTX gene AE3' age
c 62	54	0.2	38059	22	AA654018 Human factor IX (h
c 63	53	0.2	72	16	AA625953 Human gene signatu
c 64	53	0.2	192	21	AA630074 Human secreted pro
c 65	53	0.2	300	20	AA213926 Human gene express
c 66	53	0.2	350	22	AA65526 Novel human polynu
c 67	53	0.2	352	22	AA65527 Novel human polynu
c 68	53	0.2	541	20	AA679050 Human secreted pro
c 69	53	0.2	651	21	AA644206 Human secreted exp
c 70	53	0.2	746	20	AA699010 Human validated ca
c 71	53	0.2	788	21	AA674381 Human secreted pro
c 72	53	0.2	1024	21	AA681263 Human tyrosine pho
c 73	53	0.2	1721	21	AA696485 cDNA encoding a hu
c 74	53	0.2	1773	22	AA693811 Human cDNA encodin
c 75	53	0.2	1991	22	AA688096 Human FLEXHT-27 nu
c 76	53	0.2	2127	20	AA58811 Human leukaemia-as
c 77	53	0.2	2596	20	AA624899 Human secreted pro
c 78	53	0.2	5996	21	AA621203 Human low adenosin
c 79	53	0.2	6056	21	AA635081 Human adenosine re
c 80	53	0.2	7120	22	AA691210 Human folate recep
c 81	53	0.2	11811	20	AA683943 Bacterial artifi
c 82	53	0.2	13865	19	AA640401 Human tissue facto
c 83	53	0.2	13865	20	AA632165 Human cholesteryl
c 84	53	0.2	17590	21	AA650904 Human TBC-1 partia

c	85	53	0.2	22081	22	AAF97862	Human neuroblastom
86	53	0.2	24025	17	AAI17455	Mutated BRCA1 geno	
87	53	0.2	24025	17	AAI17515	Mutated BRCA1 geno	
88	53	0.2	24026	17	AAI32612	BRCA1, human breas	
89	53	0.2	24026	17	AAI17512	Mutated BRCA1 geno	
90	53	0.2	24026	17	AAI17513	Mutated BRCA1 geno	
91	53	0.2	24026	17	AAI17514	Mutated BRCA1 geno	
92	53	0.2	24026	17	AAI17516	Mutated BRCA1 geno	
93	53	0.2	24026	17	AAI17517	Mutated BRCA1 geno	
94	53	0.2	24026	17	AAI17518	Mutated BRCA1 geno	
95	53	0.2	24026	17	AAI17519	Mutated BRCA1 geno	
96	53	0.2	24026	17	AAI17521	Mutated BRCA1 geno	
97	53	0.2	24026	17	AAI17522	Mutated BRCA1 geno	
98	53	0.2	24026	17	AAI17523	Mutated BRCA1 geno	
99	53	0.2	24026	17	AAI17524	Mutated BRCA1 geno	
100	53	0.2	24026	17	AAI17526	Mutated BRCA1 geno	

ALIGNMENTS

RESULT	1	
AAA60207		
ID	AAA60207 standard; DNA; 26664 BP.	
XX		
XX	AAA60207;	
XX		
DT	07-DEC-2000 (first entry)	
DT		
XX		
DE	Human prostate cancer predisposing gene HPC2 genomic sequence.	
XX		
KW	Human; prostate cancer predisposing gene; HPC2; chromosome 17p;	
KW	gene therapy; peptide therapy; drug design; ds.	
XX		
XX		
OS	Homo sapiens.	
XX		
PH	Location/Qualifiers	
Key	910..26039	
CDS	/*tag= a	
FT	/product= "HPC2"	
FT	/note= "this sequence contains introns"	
FT	/transl_except= (pos:23892..23895,aa:Glu)	
FT		

FT		/*tag= l	
FT		/number= 11	
FT	exon	15283..15378	
FT		/*tag= m	
FT		/number= 12	
FT	exon	16278..16416	
FT		/*tag= n	
FT		/number= 13	
FT	exon	16498..16583	
FT		/*tag= o	
FT		/number= 14	
FT	exon	18583..18701	
FT		/*tag= p	
FT		/number= 15	
FT	exon	20349..20445	
FT		/*tag= q	
FT		/number= 16	
FT	exon	22172..22310	
FT		/*tag= r	
FT		/number= 17	
FT	exon	22879..22917	
FT		/*tag= s	
FT		/number= 18	
FT	exon	23045..23154	
FT		/*tag= t	
FT		/number= 19	
FT	exon	23795..23895	
FT		/*tag= u	
FT		/number= 20	
FT	exon	23973..24093	
FT		/*tag= v	
FT		/number= 21	
FT	exon	24354..24432	
FT		/*tag= w	
FT		/number= 22	
FT	exon	25026..25170	
FT		/*tag= x	
FT		/number= 23	
FT	exon	25812..26036	
FT		/*tag= y	
FT		/number= 24	
FT	polyA_signal	26447..26452	
FT		/*tag= z	
XX			
XX	WO200027864-A1.		
PN			
XX			
PD	18-MAY-2000.		
XX			
XX		99WO-US26055.	
PF	05-NOV-1999;		
XX			
XX		98US-0107468.	
PR	06-NOV-1998;		
XX			
XX	(MYRI-) MYRIAD GENETICS INC.		
PA			
XX			
PI	Tavtigian SV, Teng DHF, Simard J, Rommens JM;		
XX			
DR	WPI; 2000-376481/32.		
DR	P-PSDB; AAB07228.		
XX			
PT	Human prostate cancer (HPC)2 nucleic acids, polypeptides, and		
PT	antibodies, useful for treatment and diagnosis of prostate cancer		
XX			
XX	Claim 3; Page 108-122; 157pp; English.		
PS			
XX			
CC	The present sequence is the genomic sequence of the human prostate		
CC	cancer predisposing gene HPC2, which is found on chromosome 17p. Som		
CC	alleles of this gene cause a predisposition to cancer, particularly		
CC	prostate cancer. This gene and its protein can be used in peptide an		
CC	gene therapy for cancer patients, as well as being useful as diagnos		
CC	tools (both for cancer sufferers and those with a predisposition to		
CC	disease) and in the production of cancer drugs. This sequence was		
CC	isolated by cloning and sequencing the region of the genome whic		
CC	appeared to cause a predisposition to prostate cancer.		

Db	2041	ggattttaacccgctttaccatttttccaagtctggggtggcgagctactttttttttt	2100
Qy	2101	ttttttttttgtcagtgggcgtgatcttggctcactgcaacctttgcctctgggctcag	2160
Db	2101	ttttttttttgcagtgggcgtgatcttggctcactgcaacctttgcctctgggctcag	2160
Qy	2161	gtgacccctcacctcagcctcccaaatagtctgggacacacgctgtgcccatcacacct	2220
Db	2161	gtgacccctcacctcagcctcccaaatagtctgggacacacgctgtgcccatcacacct	2220
Qy	2221	ggctaaattttttgtattttgttagcagcgggtttttgtctatgtgtcccagctggtc	2280
Db	2221	ggctaaattttttgtattttgttagcagcgggtttttgtctatgtgtcccagctggtc	2280
Qy	2281	tcaaaacttcgcgatacctcctgtctcggcctcccagagtctgggattacaagcagatgagc	2340
Db	2281	tcaaaacttcgcgatacctcctgtctcggcctcccagagtctgggattacaagcagatgagc	2340
Qy	2341	cacgcacactggcctggaaattttttataccagcccagtcagcagcacagcagcatt	2400
Db	2341	cacgcacactggcctggaaattttttataccagcccagtcagcagcacagcagcatt	2400
Qy	2401	aaaagctgtgactcaggagaaacagattttaatatgatacacacctcttaagtgttaacct	2460
Db	2401	aaaagctgtgactcaggagaaacagattttaatatgatacacacctcttaagtgttaacct	2460
Qy	2461	ccacttagttcttcgcttgcggggacagagatttctggcagtaaaacttgagagctctagc	2520
Db	2461	ccacttagttcttcgcttgcggggacagagatttctggcagtaaaacttgagagctctagc	2520
Qy	2521	agtggtgattacagttaatatgttttacgcgacagcccattggcacattggcgaccacaca	2580
Db	2521	agtggtgattacagttaatatgttttacgcgacagcccattggcacattggcgaccacaca	2580
Qy	2581	cataccactgtccagattaccctgtcatattatgtctataccacgggaagtgccagatttgt	2640
Db	2581	cataccactgtccagattaccctgtcatattatgtctataccacgggaagtgccagatttgt	2640
Qy	2641	gttcgagcacaattgtcgtggcttggtaggcatacgaccggaagagtgaaagtgttagacca	2700
Db	2641	gttcgagcacaattgtcgtggcttggtaggcatacgaccggaagagtgaaagtgttagacca	2700
Qy	2701	gtctccctctcaggctgagactagggttgagcacttagggtgccagcccttcacttgc	2760
Db	2701	gtctccctctcaggctgagactagggttgagcacttagggtgccagcccttcacttgc	2760
Qy	2761	atgattccttaacatttgcacactgggtgccttgcgttcacacctagtgcagagctcagc	2820
Db	2761	atgattccttaacatttgcacactgggtgccttgcgttcacacctagtgcagagctcagc	2820
Qy	2821	ccattctagagcatttaagaataattggctctgtctgtttacacctctagctggcactcatt	2880
Db	2821	ccattctagagcatttaagaataattggctctgtctgtttacacctctagctggcactcatt	2880
Qy	2881	ctgcctctgtacatcttcccttgcgtgtacttccaaagctgggaaggtgagagatgtagataaa	2940
Db	2881	ctgcctctgtacatcttcccttgcgtgtacttccaaagctgggaaggtgagagatgtagataaa	2940
Qy	2941	tagttggattagtagtacgggtgcctcctctgttagtgacgacagatgcataattgatgagaga	3000
Db	2941	tagttggattagtagtacgggtgcctcctctgttagtgacgacagatgcataattgatgagaga	3000
Qy	3001	tcgtattttatgcataccttttttagaataatttacttttaaaagaaacgggcttccaa	3060
Db	3001	tcgtattttatgcataccttttttagaataatttacttttaaaagaaacgggcttccaa	3060
Qy	3061	agtggtgactttctgcacctccacaactggtagtcttctctgcacatctttcaaaaagc	3120
Db	3061	agtggtgactttctgcacctccacaacggtagtcttctctgcacatctttcaaaaagc	3120
Qy	3121	aatctttctttgttaataatcagtaaacagaattttctctttttgtcaaatcagttctctgc	3180
Db	3121	aatctttctttgttaataatcagtaaacagaattttctctttttgtcaaatcagttctctgc	3180

QY	3181	cctcagagatacctggtgctgtgaaacgcttcccccttccaagttaaagaacttgagttc	3240
Db	3181	cctcagagatacctggtgctgtgaaacgcttcccccttccaagttaaagaacttgagttc	3240
QY	3241	tgattaaactatgtgaccttgatcaagttacttttaacctttctgagcttttagttttatcctc	3300
Db	3241	tgattaaactatgtgaccttgatcaagttacttttaacctttctgagcttttagttttatcctc	3300
QY	3301	tataagatgactatcacgcttttcatagagtggttaaagattaagatgaatgacgtagcagcacata	3360
Db	3301	tataagatgactatcacgcttttcatagagtggttaaagattaagatgaatgacgtagcagcacata	3360
QY	3361	taaaagcactaaatcactttatttagatatattgtttggcaccagtaggcacacaagaagagg	3420
Db	3361	taaaagcactaaatcactttatttagatatattgtttggcaccagtaggcacacaagaagagg	3420
QY	3421	gcagctttgtttttatttatacaataaatttttgacatatcttaccctttcagtcagagtttat	3480
Db	3421	gcagctttgtttttatttatacaataaatttttgacatatcttaccctttcagtcagagtttat	3480
QY	3481	tacactcttgagaagcgctgtgtgtgtgttgaataataacagcttcattttccagtccttta	3540
Db	3481	tacactcttgagaagcgctgtgtgtgtgttgaataataacagcttcattttccagtccttta	3540
QY	3541	agaagaagtcaccaagacctgttaagcttttccccaaaataaacggtttgaatccatcca	3600
Db	3541	agaagaagtcaccaagacctgttaagcttttccccaaaataaacggtttgaatccatcca	3600
QY	3601	tttgtctcttatttgaggccttcccttatttctgttttcttatgctgtaaactaacaatagcc	3660
Db	3601	tttgtctcttatttgaggccttcccttatttctgttttcttatgctgtaaactaacaatagcc	3660
QY	3661	tcccatattcattctcgcccttccctgtaatccaatcgccacacagcagcagagagtgca	3720
Db	3661	tcccatattcattctcgcccttccctgtaatccaatcgccacacagcagcagagagtgca	3720
QY	3721	ttcaagacagaaaaagtagtgtgtcacttgcacacctaaagcccttcattgggtctcccatt	3780
Db	3721	ttcaagacagaaaaagtagtgtgtcacttgcacacctaaagcccttcattgggtctcccatt	3780
QY	3781	gcaatacaatacaaaaacaccttgatggctacaaagctctgtaggccccggccgctaccc	3840
Db	3781	gcaatacaatacaaaaacaccttgatggctacaaagctctgtaggccccggccgctaccc	3840
QY	3841	acacttccatctgtaccocatcgctgaaactgcagctgcattgggctgaactcttatgtccctc	3900
Db	3841	acacttccatctgtaccocatcgctgaaactgcagctgcattgggctgaactcttatgtccctc	3900
QY	3901	taactccctgccacttcagagacttcgcccttcgcgcgggttccctctgcctctctcta	3960
Db	3901	taactccctgccacttcagagacttcgcccttcgcgcgggttccctctgcctctctcta	3960
QY	3961	tgtcgtccctattgttactgaaccttcagggtcgcagctagagaggttcatttactccagaac	4020
Db	3961	tgtcgtccctattgttactgaaccttcagggtcgcagctagagaggttcatttactccagaac	4020
QY	4021	tgcctctcttctctctagacaagttgggataccagaactctgttattttttacttttcccttga	4080
Db	4021	tgcctctcttctctctagacaagttgggataccagaactctgttattttttacttttcccttga	4080
QY	4081	gagcacttagcataatgcactaaagctgtttctgttatcggttttccctttgtctcctcc	4140
Db	4081	gagcacttagcataatgcactaaagctgtttctgttatcggttttccctttgtctcctcc	4140
QY	4141	actggcctgattagacaagcgctccactctctttttctgtctatatccttggcactctgat	4200
Db	4141	actggcctgattagacaagcgctccactctctttttctgtctatatccttggcactctgat	4200
QY	4201	ataatgatactcagtaaatatttgcataaataatgtttcaaaaattttactaaagctttg	4260
Db	4201	ataatgatactcagtaaatatttgcataaataatgtttcaaaaattttactaaagctttg	4260

QY 4261 tttatgttgatacctattggttaaccccttttaataacttgaaatagttgctgtgtctacat 4320
DB tttatgttgatacctattggttaaccccttttaataacttgaaatagttgctgtgtctacat 4320
QY 4321 ttgttcaacataactgctcattctcttttcttattagggaaaaatacctcgaagcaat 4380
DB ttgttcaacataactgctcattctcttttcttattagggaaaaatacctcgaagcaat 4380
QY 4381 caaaatattttctgtccattgaaaggaatagaactgggtgactgctctttgtgactca 4440
DB caaaatattttctgtccattgaaaggaatagaactgggtgactgctctttgtgactca 4440
QY 4441 tctctgtctatttctaacttatatatgcctcgaacctctcaaatagaatccattaaaaac 4500
DB tctctgtctatttctaacttatatatgcctcgaacctctcaaatagaatccattaaaaac 4500
QY 4501 atcaacatacaaacctcaaaaatcaaaatgcttcatccaccagagattttttttttttt 4560
DB atcaacatacaaacctcaaaaatcaaaatgcttcatccaccagagattttttttttttttt 4560
QY 4561 ttttttgataagtcctgtctttctaccaggctggagtgagtgccatgactgcggctc 4620
DB ttttttgataagtcctgtctttctaccaggctggagtgagtgccatgactgcggctc 4620
QY 4621 actgcaacctccacccctcctgggttcaagccattctccaccctcagcctcctgagtacta 4680
DB actgcaacctccacccctcctgggttcaagccattctccaccctcagcctcctgagtacta 4680
QY 4681 ggactacaggcgcgatgccatcacgcctcagcctaatttttttatttttagtagagacgggg 4740
DB ggactacaggcgcgatgccatcacgcctcagcctaatttttttatttttagtagagacgggg 4740
QY 4741 tttcacatgttggccagatagctcgcactcttgacacttgatcctgcgcctcagc 4800
DB tttcacatgttggccagatagctcgcactcttgacacttgatcctgcgcctcagc 4800
QY 4801 ctcccaaaatgagctaccatgtggctggagatgggatttctaaatgagtacatcttctgt 4860
DB ctcccaaaatgagctaccatgtggctggagatgggatttctaaatgagtacatcttctgt 4860
QY 4861 gttcccaacctcatgctgtataaaatagggccaggctcggcaggagtgattgaacagctgat 4920
DB gttcccaacctcatgctgtataaaatagggccaggctcggcaggagtgattgaacagctgat 4920
QY 4921 gctgcctgtgtacatgctgtgtggcattctccatccacacgagcctcctgcctcag 4980
DB gctgcctgtgtacatgctgtgtggcattctccatccacacgagcctcctgcctcag 4980
QY 4981 ttccagaggtgctctcgtcgttgagttgctttgagttggggcgggggtgacaaaggtt 5040
DB ttccagaggtgctctcgtcgttgagttgctttgagttggggcgggggtgacaaaggtt 5040
QY 5041 ccttagaggttttggccaaactttgacatgaaacgagctccagctgcgcagggggg 5100
DB ccttagaggttttggccaaactttgacatgaaacgagctccagctgcgcagggggg 5100
QY 5101 cttaacagccttgatgggaagaggcctcactgagatgctagtaggctcttgcctcgg 5160
DB cttaacagccttgatgggaagaggcctcactgagatgctagtaggctcttgcctcgg 5160
QY 5161 cactggtgtgatactgtggctgttaataactcctcttttatagaacactaaactctgt 5220
DB cactggtgtgatactgtggctgttaataactcctcttttatagaacactaaactctgt 5220
QY 5221 ttcaaaataacatacagctctcttctgtgttcgatagataggttccctggcttcaactattct 5280
DB ttcaaaataacatacagctctcttctgtgttcgatagataggttccctggcttcaactattct 5280
QY 5281 gtttgttaacttgggtctctgaaagttagtactagttcttctgttttcaaattttaacg 5340
DB gtttgttaacttgggtctctgaaagttagtactagttcttctgttttcaaattttaacg 5340
QY 5341 gatgtcaccagaagattataatgtcttttcatctcgtgctgtagtaaaataggtgac 5400

DB 5341 gatgtcaccagaagattataatgtcttttcatctggtctgtagtaaaataaaatggctgac 5400
QY 5401 caaaatacaactttttatttctctaaataatgtaatactccttagaagaatcgtggttttcg 5460
DB caaaatacaactttttatttctctaaataatgtaatactccttagaagaatcgtggttttcg 5460
QY 5461 tgtcagattccacacataattctgaggaattcagttgctgctggttgggtgactcctgaag 5520
DB tgtcagattccacacataattctgaggaattcagttgctgctggttgggtgactcctgaag 5520
QY 5521 ttacttctccacacatcttctactaataatgcaatacatttgcctgtgtggttcttctgta 5580
DB ttacttctccacacatcttctactaataatgcaatacatttgcctgtgtggttcttctgta 5580
QY 5581 gctgtggtggcccccactctcctccacagaatacagagatgaaacacatgacagtttaccagatc 5640
DB gctgtggtggcccccactctcctccacagaatacagagatgaaacacatgacagtttaccagatc 5640
QY 5641 ccatacacagtgagtgatgaaagccaggtttccacagagagaggtgtacgtcctgagtaa 5700
DB ccatacacagtgagtgatgaaagccaggtttccacagagagaggtgtacgtcctgagtaa 5700
QY 5701 agaaaacatggatgaaaaatagaaaactgaaacacttgctgtgggcaacctgttttctgttct 5760
DB agaaaacatggatgaaaaatagaaaactgaaacacttgctgtgggcaacctgttttctgttct 5760
QY 5761 gagcatgattagaaaaatttagttgaggaatgaagatataggctcctcctcctggttataaa 5820
DB gagcatgattagaaaaatttagttgaggaatgaagatataggctcctcctcctggttataaa 5820
QY 5821 cttacggatgctcgtacttatgctcaatgatagtgattatgcttggaaatattagataatc 5880
DB cttacggatgctcgtacttatgctcaatgatagtgattatgcttggaaatattagataatc 5880
QY 5881 aagcacgttgtgtaaatagattgcttcaagtttgacatttgcacattcattgctggaggttttt 5940
DB aagcacgttgtgtaaatagattgcttcaagtttgacatttgcacattcattgctggaggttttt 5940
QY 5941 cccacaggcgttaataccctcttttgcatacagacatcatgaagaggtttgcacagatagat 6000
DB cccacaggcgttaataccctcttttgcatacagacatcatgaagaggtttgcacagatagat 6000
QY 6001 tttttcaaaataaataatgattacacagcaacctaaagaagtggtgttgggggttagaagct 6060
DB tttttcaaaataaataatgattacacagcaacctaaagaagtggtgttgggggttagaagct 6060
QY 6061 cctgcaaaattccgaagatctaggggccagatgatgtggtcttagcttaggaaaaagagttag 6120
DB cctgcaaaattccgaagatctaggggccagatgatgtggtcttagcttaggaaaaagagttag 6120
QY 6121 tcttgccttgaaacttgggttaaaagacattcatgtctgggttttacttactatgtgaagagag 6180
DB tcttgccttgaaacttgggttaaaagacattcatgtctgggttttacttactatgtgaagagag 6180
QY 6181 taccagcagtaggggtatttctctgttagtactaaatgtgacttactcaagttagt 6240
DB taccagcagtaggggtatttctctgttagtactaaatgtgacttactcaagttagt 6240
QY 6241 gctgtaggggtgacacaccagagcaccacagcaaaaggccagagaagtcacagaacctggcgag 6300
DB gctgtaggggtgacacaccagagcaccacagcaaaaggccagagaagtcacagaacctggcgag 6300
QY 6301 gagatgagccttacactgactgaagcagaagcagcagggagaggaatgtgcggga 6360
DB gagatgagccttacactgactgaagcagaagcagcagggagaggaatgtgcggga 6360
QY 6361 gcaatggcacaagtgtccttagccagctgtgtgagtgagctgacagcctccattgccc 6420
DB gcaatggcacaagtgtccttagccagctgtgtgagtgagctgacagcctccattgccc 6420
QY 6421 tggcttgcctcctcctcagatgctcttctcacctgacctgctgttagccaccccca 6480
DB tggcttgcctcctcctcagatgctcttctcacctgacctgctgttagccaccccca 6480

Db	10801	gctaattttt	gtattgttag	tagacacagg	gttgcgactgtg	gccaggtctag	ctctoga	10861				
Qy	10861	actcctgac	ctcaagtgat	ctgccttcct	ctcagcctcct	taagtgctggg	ttacaggagt	10920				
Db	10861	actcctgac	ctcaagtgat	ctgccttcct	ctcagcctcct	taagtgctggg	ttacaggagt	10920				
Qy	10921	gaacacat	ctccagcct	ccaattac	ctttttatt	ttattttatt	ttattttttg	10980				
Db	10921	gaacacat	ctccagcct	ccaattac	ctttttatt	ttattttatt	ttattttttg	10980				
Qy	10981	agacggagt	ctttctgtgt	gtgccagg	ctggagtg	cggtggcg	caactcttag	ctcactgc	11040			
Db	10981	agacggagt	ctttctgtgt	gtgccagg	ctggagtg	cggtggcg	caactcttag	ctcactgc	11040			
Qy	11041	aacctctc	ctccaggt	ccaagtgt	ctcatcat	cgactcc	ccgagtag	ctgggact	11100			
Db	11041	aacctctc	ctccaggt	ccaagtgt	ctcatcat	cgactcc	ccgagtag	ctgggact	11100			
Qy	11101	tcaggtgc	cccgacc	acacttgg	ctaaattt	gttttttag	tagagacg	gggtttcac	11160			
Db	11101	tcaggtgc	cccgacc	acacttgg	ctaaattt	gttttttag	tagagacg	gggtttcac	11160			
Qy	11161	catgttgc	ccaggtgt	cttgaatt	cttctga	cttctga	ctcctc	ctcctc	11220			
Db	11161	catgttgc	ccaggtgt	cttgaatt	cttctga	cttctga	ctcctc	ctcctc	11220			
Qy	11221	caaatgtc	gtggatt	acaaggt	ctgagc	ctgc	cccaac	gaagta	cttttaaacat	11280		
Db	11221	caaatgtc	gtggatt	acaaggt	ctgagc	ctgc	cccaac	gaagta	cttttaaacat	11280		
Qy	11281	tagagacat	tttagttg	ccatcct	caaaac	ccgtttgg	gtgtgg	agagaa	tgtttgggtcg	11340		
Db	11281	tagagacat	tttagttg	ccatcct	caaaac	ccgtttgg	gtgtgg	agagaa	tgtttgggtcg	11340		
Qy	11341	tgaatgtt	gttagttat	ctctaag	agatgtc	agccat	caatc	actc	actc	gtgtgtgcac	11400	
Db	11341	tgaatgtt	gttagttat	ctctaag	agatgtc	agccat	caatc	actc	actc	gtgtgtgcac	11400	
Qy	11401	actgaag	ctgtaact	ctcctcat	ctaggat	gtatttt	taagat	ggaaa	ttctaca	cccc	11460	
Db	11401	actgaag	ctgtaact	ctcctcat	ctaggat	gtatttt	taagat	ggaaa	ttctaca	cccc	11460	
Qy	11461	tgaataag	gatttc	cagatcc	aaatttg	agactc	agccct	acagta	taactct	tttaactt	11520	
Db	11461	tgaataag	gatttc	cagatcc	aaatttg	agactc	agccct	acagta	taactct	tttaactt	11520	
Qy	11521	cagagagt	taaaaga	agatgc	acagtg	tga	tgaaga	atttaa	aggagaaa	ttggaatcaaa	11580	
Db	11521	cagagagt	taaaaga	agatgc	acagtg	tga	tgaaga	atttaa	aggagaaa	ttggaatcaaa	11580	
Qy	11581	tgtcattt	agac	ctcaag	gctcac	atgtc	atctc	tgac	attttt	ctgtgtg	tgaaat	11640
Db	11581	tgtcattt	agac	ctcaag	gctcac	atgtc	atctc	tgac	attttt	ctgtgtg	tgaaat	11640
Qy	11641	ttttttt	ctccta	taaaat	gattgt	gaagtt	ttctgt	gtaga	attatt	gtttgc	ctttcta	11700
Db	11641	ttttttt	ctccta	taaaat	gattgt	gaagtt	ttctgt	gtaga	attatt	gtttgc	ctttcta	11700
Qy	11701	atgtaata	gacata	ttagggt	ttttttt	ttttttt	ttttt	ctttt	ctttt	ttttt	ttttt	11760
Db	11701	atgtaata	gacata	ttagggt	ttttttt	ttttttt	ttttt	ctttt	ctttt	ttttt	ttttt	11760
Qy	11761	ctctgc	cccg	caggtgc	agtg	agtg	agtg	agtg	agtg	agtg	agtg	11820
Db	11761	ctctgc	cccg	caggtgc	agtg	agtg	agtg	agtg	agtg	agtg	agtg	11820
Qy	11821	ctgggtt	ctcctc	cagcct	cccg	agtag	ctctg	gggact	acag	ggcg	cacgtc	11880
Db	11821	ctgggtt	ctcctc	cagcct	cccg	agtag	ctctg	gggact	acag	ggcg	cacgtc	11880
Qy	11881	gctaattt	ttttgt	atttttag	tagtag	acag	ggat	ta	cccg	ctttag	ccaggt	11940
Db	11881	gctaattt	ttttgt	atttttag	tagtag	acag	ggat	ta	cccg	ctttag	ccaggt	11940

QY	11941	atctcctgacctcatgattacacccgcctcgcttcccaaaagtctcgggattacaggcatg	12000
Db	11941	atctcctgacctcatgattacacccgcctcgcttcccaaaagtctcgggattacaggcatg	12000
QY	12001	agccgctgctggctattagagatttttattataatttatccaaagataaaagcag	12060
Db	12001	agccgctgctggctattagagatttttattataatttatccaaagataaaagcag	12060
QY	12061	tgacattattggcacataattgaaaaatacagagagaaataaaaaatccatcgctttt	12120
Db	12061	tgacattattggcacataattgaaaaatacagagagaaataaaaaatccatcgctttt	12120
QY	12121	gttagcctcaactgctcattgaaataattattgttacatggcaggtttgctgctgtgctc	12180
Db	12121	gttagcctcaactgctcattgaaataattattgttacatggcaggtttgctgctgtgctc	12180
QY	12181	tgttaggcaacgctcgtgacattcccttagctattaatgtaggaatgtagaatgacag	12240
Db	12181	tgttaggcaacgctcgtgacattcccttagctattaatgtaggaatgtagaatgacag	12240
QY	12241	aacagtgtttctcctcaatgatactgtgaagatattttataaccttccatatggaattaca	12300
Db	12241	aacagtgtttctcctcaatgatactgtgaagatattttataaccttccatatggaattaca	12300
QY	12301	ttttattaaattataatgagtttaagtctgggaataaaaaaacacgatttaagtcatcttg	12360
Db	12301	ttttattaaattataatgagtttaagtctgggaataaaaaaacacgatttaagtcatcttg	12360
QY	12361	gcttttagtactaaagcatttgacaataaaatgactcttcagaataatggataacctctg	12420
Db	12361	gcttttagtactaaagcatttgacaataaaatgactcttcagaataatggataacctctg	12420
QY	12421	aaagcaataaacgcattttaatgaaattgttaaggaaaaacaatacattttttattttt	12480
Db	12421	aaagcaataaacgcattttaatgaaattgttaaggaaaaacaatacattttttattttt	12480
QY	12481	tttttttagacacagacttctgctttgtgctaggtcgagtgacaaatggcgcgactctcg	12540
Db	12481	tttttttagacacagacttctgctttgtgctaggtcgagtgacaaatggcgcgactctcg	12540
QY	12541	gctcactgcacacatccgcctctgggttcaacgcatctctcgtcagctcctcaggttg	12600
Db	12541	gctcactgcacacatccgcctctgggttcaacgcatctctcgtcagctcctcaggttg	12600
QY	12601	ctgggattacaggcaactgcccacacacgctggccaaatttgtatttttagtagagacggg	12660
Db	12601	ctgggattacaggcaactgcccacacacgctggccaaatttgtatttttagtagagacggg	12660
QY	12661	gtttctccacgttggtcaggtgtgtctcaaaactcctgacctcaggtgatctgcgcgctc	12720
Db	12661	gtttctccacgttggtcaggtgtgtctcaaaactcctgacctcaggtgatctgcgcgctc	12720
QY	12721	agctccgaaagtagctgaattacaggcgtgagccacgctgcctggccacacattattatt	12780
Db	12721	agctccgaaagtagctgaattacaggcgtgagccacgctgcctggccacacattattatt	12780
QY	12781	tttttttaacttagaaaaatacacctctcaagaaaaattgattaaaccaaaccttctcatt	12840
Db	12781	tttttttaacttagaaaaatacacctctcaagaaaaattgattaaaccaaaccttctcatt	12840
QY	12841	agccctcaagatcacatctatgttctctttgttcagcgttccacttaagagagaaactt	12900
Db	12841	agccctcaagatcacatctatgttctctttgttcagcgttccacttaagagagaaactt	12900
QY	12901	cttgggtgctcaaaagcaagagataggcctcccagtgtagtctgggggttaaggctctct	12960
Db	12901	cttgggtgctcaaaagcaagagataggcctcccagtgtagtctgggggttaaggctctct	12960
QY	12961	ggggactcaactgggttacacctgtccacttaaggaaatacaattttcacagagccttgctt	13020
Db	12961	ggggactcaactgggttacacctgtccacttaaggaaatacaattttcacagagccttgctt	13020

QY 13021 cttcatttcagtgaggacagctgcacatgcctcccatcatttgctgctcaaggacgggaaa 13080
Db 13021 cttcatttcagtgaggacagctgcacatgcctcccatcatttgctgctcaaggacgggaaa 13080
QY 13081 agcatcactcatgaaggaagagaggtagagtgctggttttcttgatnagcaggttacag 13140
Db 13081 agcatcactcatgaaggaagagaggtagagtgctggttttcttgatnagcaggttacag 13140
QY 13141 gtagggtctgaaatgctgggacagagctgctctcttcagggccctacagacacacacttttg 13200
Db 13141 gtagggtctgaaatgctgggacagagctgctctcttcagggccctacagacacacacttttg 13200
QY 13201 aaggacgtgaaacagtttgacatcactcagctaaagtataaaatggcctcttttatctg 13260
Db 13201 aaggacgtgaaacagtttgacatcactcagctaaagtataaaatggcctcttttatctg 13260
QY 13261 tgtttgtccgcagatgtaaacacagctgcattccagagcatttttgtagattgtccatttagg 13320
Db 13261 tgtttgtccgcagatgtaaacacagctgcattccagagcatttttgtagattgtccatttagg 13320
QY 13321 atctagtccacgtccctcttaagggtgcatgcttttcccttggttacttgagctcagacag 13380
Db 13321 atctagtccacgtccctcttaagggtgcatgcttttcccttggttacttgagctcagacag 13380
QY 13381 tgtctaacaacagaccccatatggatgggctggggttttatggtccagaggaatgccaca 13440
Db 13381 tgtctaacaacagaccccatatggatgggctggggttttatggtccagaggaatgccaca 13440
QY 13441 gtattctatgtcaagatatatttctctgaactcttgagttctgaggacattaggaccacaga 13500
Db 13441 gtattctatgtcaagatatatttctctgaactcttgagttctgaggacattaggaccacaga 13500
QY 13501 ctgaagaaaccttaagtccaagccctcttctcagcagtgagcctgaaatgctcca 13560
Db 13501 ctgaagaaaccttaagtccaagccctcttctcagcagtgagcctgaaatgctcca 13560
QY 13561 acctgacagttacctgttttaggtatccaaaagagacagaagggtgtgatggtgatgt 13620
Db 13561 acctgacagttacctgttttaggtatccaaaagagacagaagggtgtgatggtgatgt 13620
QY 13621 gtaagtgtggtttgtgctttgttttaacctctcagctcactggatagatatgtcatgtta 13680
Db 13621 gtaagtgtggtttgtgctttgttttaacctctcagctcactggatagatatgtcatgtta 13680
QY 13681 gcagttgctctgaagcagttcagtttggtgctgagctgtgacccccaagtgggcggt 13740
Db 13681 gcagttgctctgaagcagttcagtttggtgctgagctgtgacccccaagtgggcggt 13740
QY 13741 tatttggttttgagattttggctgaagagctgtgtactctccagatccctggtgctgct 13800
Db 13741 tatttggttttgagattttggctgaagagctgtgtactctccagatccctggtgctgct 13800
QY 13801 ttttggtggtagaatgtccagatgaagagcttcaattcaaccccatctgtgagaatgccacc 13860
Db 13801 ttttggtggtagaatgtccagatgaagagcttcaattcaaccccatctgtgagaatgccacc 13860
QY 13861 tttcagaggtaaatgaggggtctctaggttgaggagaagtgagagctgaacccagccagc 13920
Db 13861 tttcagaggtaaatgaggggtctctaggttgaggagaagtgagagctgaacccagccagc 13920
QY 13921 atcgacatgggcatcttgttggaagagctgtgttctgggaagaccactatctgggttta 13980
Db 13921 atcgacatgggcatcttgttggaagagctgtgttctgggaagaccactatctgggttta 13980
QY 13981 cagttcagagggccggcactcctccttaagtcaactgtgttgtagttggtgggctcccggtgt 14040
Db 13981 cagttcagagggccggcactcctccttaagtcaactgtgttgtagttggtgggctcccggtgt 14040
QY 14041 acacagcctcaagtgaattagaaaagattgaaaactagaaaacactgaggactagaaa 14100
Db 14041 acacagcctcaagtgaattagaaaagattgaaaactagaaaacactgaggactagaaa 14100
QY 14101 ttcaactagaactcttacagctcttataccagaagaattctagaacttttttgaaattct 14160

Db 14101 ttcaactagaactcttacagctcttataccagaagaattctagaacttttttgaaattct 14160
QY 14161 aactaatgcccagattatcatttgattattttgaaactgaataattttcttccattac 14220
Db 14161 aactaatgcccagattatcatttgattattttgaaactgaataattttcttccattac 14220
QY 14221 ctgcatgaaacaaatgagggtgagtgctgagagctgctggtcaagagtcggt 14280
Db 14221 ctgcatgaaacaaatgagggtgagtgctgagagctgctggtcaagagtcggt 14280
QY 14281 gttatgggactggactcacagctgggaatctcttttgggctaactggcactctgtgttg 14340
Db 14281 gttatgggactggactcacagctgggaatctcttttgggctaactggcactctgtgttg 14340
QY 14341 tctctatcgaaagttaaccagattttgcgggttcagctttcattccagatggaatcattt 14400
Db 14341 tctctatcgaaagttaaccagattttgcgggttcagctttcattccagatggaatcattt 14400
QY 14401 gacccacctatctgagtttgaatctttcccccactcttaattggttacctgtatttttc 14460
Db 14401 gacccacctatctgagtttgaatctttcccccactcttaattggttacctgtatttttc 14460
QY 14461 ctgttctctagttgtatctatctgtatttttcaactgttttttctacttaccacaaca 14520
Db 14461 ctgttctctagttgtatctatctgtatttttcaactgttttttctacttaccacaaca 14520
QY 14521 aatccctttgggctgctgtacccctccagagtcagagcgttagaggtgtttcatgctct 14580
Db 14521 aatccctttgggctgctgtacccctccagagtcagagcgttagaggtgtttcatgctct 14580
QY 14581 gctttatctctgtgggtgaatttgatgcgtgggtagccccgcttcttattttaacc 14640
Db 14581 gctttatctctgtgggtgaatttgatgcgtgggtagccccgcttcttattttaacc 14640
QY 14641 agtttggggcagcaaaacctcttcaatgaatgaatgggtgtctatttgagagccatgtgtgat 14700
Db 14641 agtttggggcagcaaaacctcttcaatgaatgaatgggtgtctatttgagagccatgtgtgat 14700
QY 14701 gtgtgatgctgctgggagatagataaaaaatagctactgtgtattttttaaagggaac 14760
Db 14701 gtgtgatgctgctgggagatagataaaaaatagctactgtgtattttttaaagggaac 14760
QY 14761 tgagagaaacacatcagcatgttagtaagtgtctgtgtccaggtggtggaatttcag 14820
Db 14761 tgagagaaacacatcagcatgttagtaagtgtctgtgtccaggtggtggaatttcag 14820
QY 14821 atgattttcatttctcgtgctgtgtctcaggtctctcaggtggaagcagacacccaggtggc 14880
Db 14821 atgattttcatttctcgtgctgtgtctcaggtctctcaggtggaagcagacacccaggtggc 14880
QY 14881 attgaggtgacagagaggtttatttcagagaaaatttgactgtgagagaggaagagaggg 14940
Db 14881 attgaggtgacagagaggtttatttcagagaaaatttgactgtgagagaggaagagaggg 14940
QY 14941 agcagaagagcagagagagcctggtctggtttcaggtttgacccctatgagtgga 15000
Db 14941 agcagaagagcagagagagcctggtctggtttcaggtttgacccctatgagtgga 15000
QY 15001 gagggtaggaaggaagtgcagtgctgagaagagatcagccagcctacttgaaaagccag 15060
Db 15001 gagggtaggaaggaagtgcagtgctgagaagagatcagccagcctacttgaaaagccag 15060
QY 15061 agcagagcttgcagatatacaggaatcccatcgtccatttggaaatggccagcagccgggtc 15120
Db 15061 agcagagcttgcagatatacaggaatcccatcgtccatttggaaatggccagcagccgggtc 15120
QY 15121 tgcgctgagcagcctgctgtgagagcatgctggcggtgaggtctgagctcagctcactgca 15180
Db 15121 tgcgctgagcagcctgctgtgagagcatgctggcggtgaggtctgagctcagctcactgca 15180
QY 15181 gtgctgcagagggcccgacgatccccctcctgctgctgctggctgcttctgtgtgtgtc 15240

Db	15181	gtctgcagaggccgacagataccccctccctgctgctcggttcctgtcttctgtgtc	15244
Qy	15241	ctgagtctgatcaactttgtaaagcccacatctcttgcccaggtaccgaaggaagcaga	15300
Db	15241	ctgagtctgatcaactttgtaaagcccacatctctgccaggtaccaaggaagcaga	15300
Qy	15301	tgcocccgtggccttgtgtgttccacatgccccagcatctgtgttggcacagcagta	15360
Db	15301	tgcocccgtggccttgtgtgttccacatgccccagcatctgtgttggcacagcagta	15360
Qy	15361	ccacagtgtagtgagaggtatgagccacgcccagcgggcaacttgggttaactctctgg	15420
Db	15361	ccacagtgtagtgagaggtatgagccacgcccagcgggcaacttgggttaactctctgg	15420
Qy	15421	gcagtgttggaattcccctttccctccctgtgtctttccagcgtacctaccctctgc	15480
Db	15421	gcagtgttggaattcccctttccctccctgtgtctttccagcgtacctaccctctgc	15480
Qy	15481	acctgctaaaactttctgtgtggaattctgtccttcccagaaattctaggtctcccagatctg	15540
Db	15481	acctgctaaaactttctgtgtggaattctgtccttcccagaaattctaggtctcccagatctg	15540
Qy	15541	tgtctacactgtgaagaaatgcacgcctaagtggcgacgtgtccacacgattccattta	15600
Db	15541	tgtctacactgtgaagaaatgcacgcctaagtggcgacgtgtccacacgattccattta	15600
Qy	15601	ttttacacctccacactcttcagggtgtctgacaataactgcogtttgggttgaggaatt	15660
Db	15601	ttttacacctccacactcttcagggtgtctgacaataactgcogtttgggttgaggaatt	15660
Qy	15661	cctaagtgaattccaagaagagattgcagctataaaatgatagcttccatttactgaa	15720
Db	15661	cctaagtgaattccaagaagagattgcagctataaaatgatagcttccatttactgaa	15720
Qy	15721	tgccaatcttgtggaggcaggtgtgaaatacoccctcaatttcaacttctctctagg	15780
Db	15721	tgccaatcttgtggaggcaggtgtgaaatacoccctcaatttcaacttctctctagg	15780
Qy	15781	gtcgtcgccagcagccctggaggtagatgttttagtcaactggaagcatcttttctctcg	15840
Db	15781	gtcgtcgccagcagccctggaggtagatgttttagtcaactggaagcatcttttctctcg	15840
Qy	15841	gggcatcgtctggccaggccaggttggaggagtagttagtgcagctcgggtgcgggtgacc	15900
Db	15841	gggcatcgtctggccaggccaggttggaggagtagttagtgcagctcgggtgcgggtgacc	15900
Qy	15901	tgggctgcgttttttggccctgcgcgatatctcccacatggcccggtttaactgcacctca	15960
Db	15901	tgggctgcgttttttggccctgcgcgatatctcccacatggcccggtttaactgcacctca	15960
Qy	15961	ctccatggctgcctcctcgtctctcttcttcaactcaggggttgggtcccttatitgt	16020
Db	15961	ctccatggctgcctcctcgtctctcttcttcaactcaggggttgggtcccttatitgt	16020
Qy	16021	atgcccctggacatcttctttctgttttctctcactcttcccagcacacctgaaag	16080
Db	16021	atgcccctggacatcttctttctgttttctctcactcttcccagcacacctgaaag	16080
Qy	16081	gcagctgagctagggaacacccggctttgagacagcaggagtggaacatgtttggccat	16140
Db	16081	gcagctgagctagggaacacccggctttgagacagcaggagtggaacatgtttggccat	16140
Qy	16141	gtagtaaacctgcttgggcaagtcactgaacttttgaacacctcatctcataccac	16200
Db	16141	gtagtaaacctgcttgggcaagtcactgaacttttgaacacctcatctcataccac	16200
Qy	16201	tccttgactcagcaccactcctcaggggagctgccttccctaaacagcgtcgaaatgccg	16260
Db	16201	tccttgactcagcaccactcctcaggggagctgccttccctaaacagcgtcgaaatgccg	16260
Qy	16261	ggctgtttcttcacaggtttggccctgaacaccagcacttggctcctgaaatggaactgt	16320
Db	16261	ggctgtttcttcacaggtttggccctgaacaccagcacttggctcctgaaatggaactgt	16320

QY	16321	gctcagttccaacaccttcgcgaagccacaagattcaaaaccacgctcaaacctcatcacccg	16380
Db	16321	gctcagttccaacaccttcgcgaagccacaagattcaaaaccacgctcaaacctcatcacccg	16380
QY	16381	qacatcttccccctgctcaccagtttcgcgtgaaggtagtgctcaacyggccccttg	16440
Db	16381	gacatcttccccctgctcaccagtttcgcgtgaaggtagtgctcaacyggccccttg	16440
QY	16441	tgggccagctctcgctccctctctttctccatgaatgtgtttgtctcttttcagaag	16500
Db	16441	tgggccagctctcgctccctctctttctccatgaatgtgtttgtctcttttcagaag	16500
QY	16501	gagggcccaacctcaagtgtgccatggttcagggtgaatgcctcctcaagtaccagctc	16560
Db	16501	gagggcccaacctcaagtgtgccatggttcagggtgaatgcctcctcaagtaccagctc	16560
QY	16561	cgtccagaggaggagtgccagaggtctgtgcactcttgaaactaatggaatcgtctcagtc	16620
Db	16561	cgtccagaggaggagtgccagaggtctgtgcactcttgaaactaatggaatcgtctcagtc	16620
QY	16621	gagttgggaacattcttgttaactatcaactccgctttgagcaciaaagaaaacacggacaat	16680
Db	16621	gagttgggaacattcttgttaactatcaactccgctttgagcaciaaagaaaacacggacaat	16680
QY	16681	catatggtctctgttgtaactatcaactccgctttgagcaciaaagaaaacacggacaat	16740
Db	16681	catatggtctctgttgtaactatcaactccgctttgagcaciaaagaaaacacggacaat	16740
QY	16741	aagctaaatgaatgggcttggtgtgtgccagctgaaatttatagaaaaagcgcctact	16800
Db	16741	aagctaaatgaatgggcttggtgtgtgccagctgaaatttatagaaaaagcgcctact	16800
QY	16801	ccaggtcgtggttgaggtggcggaattggggccagtagttctccttttccaaaaatgacct	16860
Db	16801	ccaggtcgtggttgaggtggcggaattggggccagtagttctccttttccaaaaatgacct	16860
QY	16861	gcatagggaatagcagtgatagagctcgtgttttcacagtagataaaaaatggaatgtgt	16920
Db	16861	gcatagggaatagcagtgatagagctcgtgttttcacagtagataaaaaatggaatgtgt	16920
QY	16921	gatgaacaaagtccaccaatactcgttgcgccagagataatgtagtaacattttgtgt	16980
Db	16921	gatgaacaaagtccaccaatactcgttgcgccagagataatgtagtaacattttgtgt	16980
QY	16981	ttcttgatttgtatgfgggtttatatattgcagctcttttccctgtatcaactaaacagctc	17040
Db	16981	ttcttgatttgtatgfgggtttatatattgcagctcttttccctgtatcaactaaacagctc	17040
QY	17041	taagtacaagaatttttatggttatccaaataggagtggtttactcatattgggagtggttc	17100
Db	17041	taagtacaagaatttttatggttatccaaataggagtggtttactcatattgggagtggttc	17100
QY	17101	caatttttgtgtttttaatgaatgaacaaataaattcttatataaacttttgatg	17160
Db	17101	caatttttgtgtttttaatgaatgaacaaataaattcttatataaacttttgatg	17160
QY	17161	ggaaactcgttccctccaagtcaattcctaattgtgggattactgcccagagtgtagac	17220
Db	17161	ggaaactcgttccctccaagtcaattcctaattgtgggattactgcccagagtgtagac	17220
QY	17221	ttgttaagggtacttgataaattgtaagatgccaattottgaaagcctcttccagtacaaacca	17280
Db	17221	ttgttaagggtacttgataaattgtaagatgccaattottgaaagcctcttccagtacaaacca	17280
QY	17281	accaggaaagtgaacagccttactgccccacatcttatttttaattaataattatttta	17340
Db	17281	accaggaaagtgaacagccttactgccccacatcttatttttaattaataattatttta	17340
QY	17341	ttttatttattattatttttttgagacggagtttctactctgtgtcccaggctggaggt	17400
Db	17341	ttttatttattattatttttttgagacggagtttctactctgtgtcccaggctggaggt	17400

Db 19561 tggcttcgagtgagagctggatgcaaggccacacactgttctcttggaaaaatgacagt 19620
Qy 19621 tccagcataaagaaatgctcatcttttaaatgtataaagaattacagcaattcttttgaag 19680
Db 19621 tccagcataaagaaatgctcatcttttaaatgtataaagaattacagcaattcttttgaag 19680
Qy 19681 aaaggactggagaatttatcttcttcttagccttttggtagacagatagccttgggctc 19740
Db 19681 aaaggactggagaatttatcttcttcttagccttttggtagacagatagccttgggctc 19740
Qy 19741 ccacactggctgcgaagtcctttgtttcagagcggttgccaggggcgctgccagtcceccctc 19800
Db 19741 ccacactggctgcgaagtcctttgtttcagagcggttgccaggggcgctgccagtcceccctc 19800
Qy 19801 ctgggaagctggatagaactatgttgttaccatctgtctttagctgtgttttggttatt 19860
Db 19801 ctgggaagctggatagaactatgttgttaccatctgtctttagctgtgttttggttatt 19860
Qy 19861 ataaaaaatatgtgagactgggttaattatcaagaaaaagaggtgtatttagttcacgggt 19920
Db 19861 ataaaaaatatgtgagactgggttaattatcaagaaaaagaggtgtatttagttcacgggt 19920
Qy 19921 tctgcgggctgagaattgaaaggtccagcgccctagcttccagtgaaaggtctccatgtcgca 19980
Db 19921 tctgcgggctgagaattgaaaggtccagcgccctagcttccagtgaaaggtctccatgtcgca 19980
Qy 19981 tcataacgtgcagagagacgcaagtaggaagtggacgcttggaagacgggataaacctg. 20040
Db 19981 tcataacgtgcagagagacgcaagtaggaagtggacgcttggaagacgggataaacctg 20040
Qy 20041 agctgcactctggttttataaacacccctctctctggtggacaaaaccattcccttgaga 20100
Db 20041 agctgcactctggttttataaacacccctctctctggtggacaaaaccattcccttgaga 20100
Qy 20101 agtaatgcagctctcctgagagccagtgacttacttactgagctcagctccaaagccactcaggagg 20160
Db 20101 agtaatgcagctctcctgagagccagtgacttacttactgagctcagctccaaagccactcaggagg 20160
Qy 20161 tccgtccctgtagccaaaacgcttccactagagcccgctcccaaaacggcatactag 20220
Db 20161 tccgtccctgtagccaaaacgcttccactagagcccgctcccaaaacggcatactag 20220
Qy 20221 ggagcacggtttccacatgaggtctgggacaaaaccaatgacactcaaacattgcaacctt 20280
Db 20221 ggagcacggtttccacatgaggtctgggacaaaaccaatgacactcaaacattgcaacctt 20280
Qy 20281 ctcatggctgcatgctggctcaacttttgacccaaaggaatggattgttccatggatt 20340
Db 20281 ctcatggctgcatgctggctcaacttttgacccaaaggaatggattgttccatggatt 20340
Qy 20341 ttccacagagaaaagagtcagtcaccagaaaacatctctcttggaaacagggtctgccat 20400
Db 20341 ttccacagagaaaagagtcagtcaccagaaaacatctctcttggaaacagggtctgccat 20400
Qy 20401 cccgataagatttcgaatgcagtcgcaacttgcacataaagtatgctgctttccc 20460
Db 20401 cccgataagatttcgaatgcagtcgcaacttgcacataaagtatgctgctttccc 20460
Qy 20461 aggaagcatccttccatcaagggcaygtttactttttaaacaagaatcctgcgtactca 20520
Db 20461 aggaagcatccttccatcaagggcaygtttactttttaaacaagaatcctgcgtactca 20520
Qy 20521 ccagtcgatttgaatgcggtatcaagccctgtcacctgtcatgtcgactggagtgcca 20580
Db 20521 ccagtcgatttgaatgcggtatcaagccctgtcacctgtcatgtcgactggagtgcca 20580
Qy 20581 ggagagagcgtggccttactgcattttagcctcagtagcagaaaactttaccctgggaat 20640
Db 20581 ggagagagcgtggccttactgcattttagcctcagtagcagaaaactttaccctgggaat 20640
Qy 20641 caccaaaattcattcccatgatgctttttaaataaacagctgattttactgtgggcagtaca 20700
Db 20641 caccaaaattcattcccatgatgctttttaaataaacagctgattttactgtgggcagtaca 20700

Qy 20701 cctagctaaagaaattagctcctttaatttttacattaatcctatgaagtgtgtaataact 20760
Db 20701 cctagctaaagaaattagctcctttaatttttacattaatcctatgaagtgtgtaataact 20760
Qy 20761 accattttgtgatgagtgacctgatatccagagaggtgacttgctatggttccctacag 20820
Db 20761 accattttgtgatgagtgacctgatatccagagaggtgacttgctatggttccctacag 20820
Qy 20821 ctggtaagtggggaactctgaagtttgagcggggacttggggtctttagtgcatacatgta 20880
Db 20821 ctggtaagtggggaactctgaagtttgagcggggacttggggtctttagtgcatacatgta 20880
Qy 20881 ttgtcccccagccatttggtagtatgttaaaaagctttagggttttgcacatttctg 20940
Db 20881 ttgtcccccagccatttggtagtatgttaaaaagctttagggttttgcacatttctg 20940
Qy 20941 ttcagaacacctttagtattcccttgacatgttttttagttgattctcttgggtttgcc 21000
Db 20941 ttcagaacacctttagtattcccttgacatgttttttagttgattctcttgggtttgcc 21000
Qy 21001 tggggtcacagcagagagatttagtcaaatcggttgacatgtacacgttatctctaca 21060
Db 21001 tggggtcacagcagagagatttagtcaaatcggttgacatgtacacgttatctctaca 21060
Qy 21061 gatagtatgtgaagaaaaataagattgtgaattaccaggtttgttttaaaatttgcctgc 21120
Db 21061 gatagtatgtgaagaaaaataagattgtgaattaccaggtttgttttaaaatttgcctgc 21120
Qy 21121 catctacatgctagtggttgatgataaacacaaatagtgattaaatatatacacga 21180
Db 21121 catctacatgctagtggttgatgataaacacaaatagtgattaaatatatacacga 21180
Qy 21181 gtacagagatgtcccttgacatcagaaaaatacaaatctggggtgtgttctctctgagatg 21240
Db 21181 gtacagagatgtcccttgacatcagaaaaatacaaatctggggtgtgttctctctgagatg 21240
Qy 21241 aggacatgcataaagacagcttgagtgagccgctctcccggggctgagatccctggg 21300
Db 21241 aggacatgcataaagacagcttgagtgagccgctctcccggggctgagatccctggg 21300
Qy 21301 ggaagaaggcctttttagtttgacctgacacctcgagcagcttttgaccagcgtgaa 21360
Db 21301 ggaagaaggcctttttagtttgacctgacacctcgagcagcttttgaccagcgtgaa 21360
Qy 21361 gctaatgggaagtgctattgcccacttgcctcgcctccgactccttttcccacaga 21420
Db 21361 gctaatgggaagtgctattgcccacttgcctcgcctccgactccttttcccacaga 21420
Qy 21421 aggtaatgtcttagcaccgggcttctctgcataatgggtgcagccctctcagtgctc 21480
Db 21421 aggtaatgtcttagcaccgggcttctctgcataatgggtgcagccctctcagtgctc 21480
Qy 21481 gtggtcctcccagagaaatgaaggagccagagcggtgcagcactctctctgcttgag 21540
Db 21481 gtggtcctcccagagaaatgaaggagccagagcggtgcagcactctctctgcttgag 21540
Qy 21541 cagagcttctgaatggactgcacagagaaatagcccaagaagtgtgtcagaatccaagac 21600
Db 21541 cagagcttctgaatggactgcacagagaaatagcccaagaagtgtgtcagaatccaagac 21600
Qy 21601 ttcagagccctgcctaaaaaccaagtgcagaaacccgagtgacacctggaggtctcgctt 21660
Db 21601 ttcagagccctgcctaaaaaccaagtgcagaaacccgagtgacacctggaggtctcgctt 21660
Qy 21661 aactggtcctcctgaatgaagcacctgcagcccgccctgcaccagggtgtctttgaggacat 21720
Db 21661 aactggtcctcctgaatgaagcacctgcagcccgccctgcaccagggtgtctttgaggacat 21720
Qy 21721 gactgaggaaccccgaccacttgcaaaaggggaaaaagtcacgatggcagctggaccctag 21780
Db 21721 gactgaggaaccccgaccacttgcaaaaggggaaaaagtcacgatggcagctggaccctag 21780

QY 21781 aaagagtctcatatggccagctgcctgtctctgtatttcaacagagagctgtggccacag 21840
|||||
Db 21781 aaagagtctcatatggccagctgcctgtctctgtatttcaacagagagctgtggccacag 21840
|||||
QY 21841 tcaatctgcattgctcagattcatatttaggactaaatgctttaaagccctcctataaaccttt 21900
|||||
Db 21841 tcaatctgcattgctcagattcatatttaggactaaatgctttaaagccctcctataaaccttt 21900
|||||
QY 21901 ttttttttttttggatgccagcctttgttaagtctacttgaaagggtttcagggttc 21960
|||||
Db 21901 ttttttttttttggatgccagcctttgttaagtctacttgaaagggtttcagggttc 21960
|||||
QY 21961 catgataactctttgctataaagagagatgacacatgtaaaatcacctttatgggttaaa 22020
|||||
Db 21961 catgataactctttgctataaagagagatgacacatgtaaaatcacctttatgggttaaa 22020
|||||
QY 22021 taattggcttttatattagctcctcaagcaaaagcaggagagacagaataattctcagtt 22080
|||||
Db 22021 taattggcttttatattagctcctcaagcaaaagcaggagagacagaataattctcagtt 22080
|||||
QY 22081 gctcttggtctgtccaaagcagacatcagcctctgaacctatcagcagctctctctagtg 22140
|||||
Db 22081 gctcttggtctgtccaaagcagacatcagcctctgaacctatcagcagctctctctagtg 22140
|||||
QY 22141 gcatgactctctctctctctctctgcagcccgacacgtctctgctactggactgtg 22200
|||||
Db 22141 gcatgactctctctctctctctctgcagcccgacacgtctctgctactggactgtg 22200
|||||
QY 22201 gtgagggcaactttggcagctgtgccgtcattacggagacaggtggacagggctcctgg 22260
|||||
Db 22201 gtgagggcaactttggcagctgtgccgtcattacggagacaggtggacagggctcctgg 22260
|||||
QY 22261 gcaacctggctgtgtttgtgtccacctgcacgcagatcaccacacggtgagtggtg 22320
|||||
Db 22261 gcaacctggctgtgtttgtgtccacctgcacgcagatcaccacacggtgagtggtg 22320
|||||
QY 22321 ggtcgacacaaaagctggagcctggagagcactgcacgttgagttggccctttgagc 22380
|||||
Db 22321 ggtcgacacaaaagctggagcctggagagcactgcacgttgagttggccctttgagc 22380
|||||
QY 22381 tgcgtctttctcgccttccaaacttgcacagactttgttaactcatctctgctagg 22440
|||||
Db 22381 tgcgtctttctcgccttccaaacttgcacagactttgttaactcatctctgctagg 22440
|||||
QY 22441 aaatggtttttgcaaaactcaacatagctctcttgcgccacaagaatgtctctctcc 22500
|||||
Db 22441 aaatggtttttgcaaaactcaacatagctctcttgcgccacaagaatgtctctctcc 22500
|||||
QY 22501 tgttcagttcctttctgcagcagacaggtttgagtttacccagcctctccttgacttt 22560
|||||
Db 22501 tgttcagttcctttctgcagcagacaggtttgagtttacccagcctctccttgacttt 22560
|||||
QY 22561 gaatctcaacggcctgtctcagcgaaagcttgaacggatgcagaggtgtggtgctatgag 22620
|||||
Db 22561 gaatctcaacggcctgtctcagcgaaagcttgaacggatgcagaggtgtggtgctatgag 22620
|||||
QY 22621 accctcaactgtctctcgtgggtaccggccctggccgttgccctcttccacagcag 22680
|||||
Db 22621 accctcaactgtctctcgtgggtaccggccctggccgttgccctcttccacagcag 22680
|||||
QY 22681 gtcgtgcgtctttctgcctgtgacatttcagggccatggcgagggggtcggcctgtgc 22740
|||||
Db 22681 gtcgtgcgtctttctgcctgtgacatttcagggccatggcgagggggtcggcctgtgc 22740
|||||
QY 22741 caccocccactgcggctgtgttagaggtgtgtgggtgaacgtcggggtcggcaactcctgcaa 22800
|||||
Db 22741 caccocccactgcggctgtgttagaggtgtgtgggtgaacgtcggggtcggcaactcctgcaa 22800
|||||
QY 22801 gagagaggtctcagaccctaaacccgagggatggccctggggcctggctgagcagcatgt 22860
|||||
Db 22801 gagagaggtctcagaccctaaacccgagggatggccctggggcctggctgagcagcatgt 22860
|||||
QY 22861 ctccctgtttccttgcagggcttgcaaatctctgtctcagagagaaacgcgcttggtta 22920
|||||

Db 22861 ctccctgtttccttgcagggcttgccaaagtatctgtctcagagagaacgcgcttggtta 22920
|||||
QY 22921 agtgtggcaacttgatggcggttctgaagtttcaagcgtttacacatcatccgcacgtctc 22980
|||||
Db 22921 agtgtggcaacttgatggcggttctgaagtttcaagcgtttacacatcatccgcacgtctc 22980
|||||
QY 22981 ttggcaactcagtttttattgagatgtctcgtcagagtcggcacttgcattttttgtt 23040
|||||
Db 22981 ttggcaactcagtttttattgagatgtctcgtcagagtcggcacttgcattttttgtt 23040
|||||
QY 23041 ccaggcatcttttggaaagccgttccacctttgctgtgtgttgggttgcoccccaacagctcaa 23100
|||||
Db 23041 ccaggcatcttttggaaagccgttccacctttgctgtgtgttgggttgcoccccaacagctcaa 23100
|||||
QY 23101 agcctggctccagcagtcaccaaacccagtcgcaagggttccctgcaccacatcaggtgagc 23160
|||||
Db 23101 agcctggctccagcagtcaccaaacccagtcgcaagggttccctgcaccacatcaggtgagc 23160
|||||
QY 23161 atccagggcagcctggcccgstgggctgtgtgtgtgcgtctccttccagaaagctcaag 23220
|||||
Db 23161 atccagggcagcctggcccgstgggctgtgtgtgtgcgtctccttccagaaagctcaag 23220
|||||
QY 23221 gtggacactggggttagttaccacatataccccccagcagccttgccttgacatggctccaga 23280
|||||
Db 23221 gtggacactggggttagttaccacatataccccccagcagccttgccttgacatggctccaga 23280
|||||
QY 23281 tggcagaagcaggggagaagtgcatttgcgtgaagcagaacacacattagatagttcccat 23340
|||||
Db 23281 tggcagaagcaggggagaagtgcatttgcgtgaagcagaacacacattagatagttcccat 23340
|||||
QY 23341 gtaatgcttattttctagaagcattttccacgttccatcttgcgtctccttgcagctgctt 23400
|||||
Db 23341 gtaatgcttattttctagaagcattttccacgttccatcttgcgtctccttgcagctgctt 23400
|||||
QY 23401 tctaaacttcgagcagcttttcttgatgagacagtttccagagccaaagcaccacaaatagtg 23460
|||||
Db 23401 tctaaacttcgagcagcttttcttgatgagacagtttccagagccaaagcaccacaaatagtg 23460
|||||
QY 23461 gtagcacagagaatgtccatagcaggtgtgtgtgtagctgtgcaggtggcacatcctca 23520
|||||
Db 23461 gtagcacagagaatgtccatagcaggtgtgtgtgtagctgtgcaggtggcacatcctca 23520
|||||
QY 23521 ccccaagggaaggaagtcctctctgcgtgagccatccgtgcgcccgtgcctgcagcccg 23580
|||||
Db 23521 ccccaagggaaggaagtcctctctgcgtgagccatccgtgcgcccgtgcctgcagcccg 23580
|||||
QY 23581 aggcagcattccactgctgggtttctccagtgccctagaggtttgggttggctcttta 23640
|||||
Db 23581 aggcagcattccactgctgggtttctccagtgccctagaggtttgggttggctcttta 23640
|||||
QY 23641 tatttgactgctgtttctctcatcatagtagtactatgatttaactcatgttttctcctaaaga 23700
|||||
Db 23641 tatttgactgctgtttctctcatcatagtagtactatgatttaactcatgttttctcctaaaga 23700
|||||
QY 23701 atgattttgggtgtctccagccaaagacttaaaccttgcgtccagatgtccaaagaacgt 23760
|||||
Db 23701 atgattttgggtgtctccagccaaagacttaaaccttgcgtccagatgtccaaagaacgt 23760
|||||
QY 23761 ttattatcatatttaaatgttttctttttacagtagtattcctgcacaaatgccttcagg 23820
|||||
Db 23761 ttattatcatatttaaatgttttctttttacagtagtattcctgcacaaatgccttcagg 23820
|||||
QY 23821 aagggtctgagatctccagctcctgcagtggaagaatgtagatcagttgcgtgtgtgcgaact 23880
|||||
Db 23821 aagggtctgagatctccagctcctgcagtggaagaatgtagatcagttgcgtgtgtgcgaact 23880
|||||
QY 23881 gtgatttgaaaggttaaggccacagccgagcatcatggggcaggtgggagagcag 23940
|||||
Db 23881 gtgatttgaaaggttaaggccacagccgagcatcatggggcaggtgggagagcag 23940
|||||
QY 23941 agctgcagagcctccagcccccaccccttccagttccagacctgtctgttgcggcactgca 24000
|||||

Db 23941 agctgcagagccctccagcccccaccctttcagtttcagaaacctgtctgtggtgcactgca 24000
QY 24001 agcatgcgtttggctgtgcgtgctggtgcacacctctgctggaagtgtctattccgggg 24060
Db 24001 agcatgcgtttggctgtgcgtgctggtgcacacctctgctggaagtgtctattccgggg 24060
QY 24061 acaccatgcacctgcagagctctggtccggatggtgagtagaggaagaagcaagccacc 24120
Db 24061 acaccatgcacctgcagagctctggtccggatggtgagtagaggaagaagcaagccacc 24120
QY 24121 tgagggtgtctctgggtttgtgagctgcaggtgaaatgcaggtgggtctcagggaaacg 24180
Db 24121 tgagggtgtctctgggtttgtgagctgcaggtgaaatgcaggtgggtctcagggaaacg 24180
QY 24181 tcagcagagcagagagactcaggtcccccacctcagagttctcgtttgtcatctcagtag 24240
Db 24181 tcagcagagcagagagactcaggtcccccacctcagagttctcgtttgtcatctcagtag 24240
QY 24241 gcagaccagggccaggggagctgagttgtgagaccaggaacacagcactgactgagccc 24300
Db 24241 gcagaccagggccaggggagctgagttgtgagaccaggaacacagcactgactgagccc 24300
QY 24301 tgtgtccgctctcgcagagaactctccctgatcctctgtgctgtctctccagggaaga 24360
Db 24301 tgtgtccgctctcgcagagaactctccctgatcctctgtgctgtctctccagggaaga 24360
QY 24361 tgcacacctctgatcatgaagccaccttgaagtgttggaaagaggaagcagtgga 24420
Db 24361 tgcacacctctgatcatgaagccaccttgaagtgttggaaagaggaagcagtgga 24420
QY 24421 aaagacacacaggtagcaaaaggccgtcagttcttgcgccacatcctctccctccccc 24480
Db 24421 aaagacacacaggtagcaaaaggccgtcagttcttgcgccacatcctctccctccccc 24480
QY 24481 actacgtgacactgagcagccgtgtgtgtctccactgagtgggggtgcctgtctcct 24540
Db 24481 actacgtgacactgagcagccgtgtgtgtctccactgagtgggggtgcctgtctcct 24540
QY 24541 atcaagggtcatgggggtccttgcactgtgcagtggtcagaggtcttgcctttat 24600
Db 24541 atcaagggtcatgggggtccttgcactgtgcagtggtcagaggtcttgcctttat 24600
QY 24601 ttttgagaattttctaagcaagatttagagttaggcacagcttttttgaagcatatag 24660
Db 24601 ttttgagaattttctaagcaagatttagagttaggcacagcttttttgaagcatatag 24660
QY 24661 aaatcggctgaataaactataagccatgtcagggaattgcagggggaagcggggctgg 24720
Db 24661 aaatcggctgaataaactataagccatgtcagggaattgcagggggaagcggggctgg 24720
QY 24721 gggactgaatttttggctgctaattcaacgaaagagtgcatlaccocaggtggccctg 24780
Db 24721 gggactgaatttttggctgctaattcaacgaaagagtgcatlaccocaggtggccctg 24780
QY 24781 tggttctcttgggtgccctcatggacagatttggcagccagcacagaggtgggcttca 24840
Db 24781 tggttctcttgggtgccctcatggacagatttggcagccagcacagaggtgggcttca 24840
QY 24841 tcagggggtgtgtgcgaagctctggccctcaggggagattgtgctggtcacaggaggtgc 24900
Db 24841 tcagggggtgtgtgcgaagctctggccctcaggggagattgtgctggtcacaggaggtgc 24900
QY 24901 ccgttaagaaaacccacagcttcccgggtgccttgcagttgagtgcagcaggtctgtg 24960
Db 24901 ccgttaagaaaacccacagcttcccgggtgccttgcagttgagtgcagcaggtctgtg 24960
QY 24961 ccactgtctgtttgcagttctgcagttgagttcagcttcagttcagttcgtctgtcctcacc 25020
Db 24961 ccactgtctgtttgcagttctgcagttgagttcagttcagttcagttcgtctgtcctcacc 25020
QY 25021 tgcagacacagtcaccaagccatcagcgtggggatgcggatgaacgcggaggtcattatg 25080
Db 25021 tgcagacacagtcaccaagccatcagcgtggggatgcggatgaacgcggaggtcattatg 25080

QY 25081 ctgaaccaacttcagccagcgctatgccaaaggtcccccctcttcagccccaacttcagcgag 25140
Db 25081 ctgaaccaacttcagccagcgctatgccaaaggtcccccctcttcagccccaacttcagcgag 25140
QY 25141 aaagtgggagttgcctttgaccacatgaaggtctgtatgtccacgcgacagcagggcg 25200
Db 25141 aaagtgggagttgcctttgaccacatgaaggtctgtatgtccacgcgacagcagggcg 25200
QY 25201 ggcagcgggagggagagacaggaactctacacactgagtaggacggtcagctggaatttgcct 25260
Db 25201 ggcagcgggagggagagacaggaactctacacactgagtaggacggtcagctggaatttgcct 25260
QY 25261 ttcttatttggggccaccgtgggaaaggttatctaccctcactaaccacaggtcgaaacca 25320
Db 25261 ttcttatttggggccaccgtgggaaaggttatctaccctcactaaccacaggtcgaaacca 25320
QY 25321 ccttggtgttgcgtggtgagaccacacctcctgcagggggccaaactagtcttcagttcagtt 25380
Db 25321 ccttggtgttgcgtggtgagaccacacctcctgcagggggccaaactagtcttcagttcagtt 25380
QY 25381 cactggaattttctgagaatcccttttaggctggaactgcacacagtcactggtcatttga 25440
Db 25381 cactggaattttctgagaatcccttttaggctggaactgcacacagtcactggtcatttga 25440
QY 25441 gctcagcacagacctgtgagacaggtggttgcctctgtgagtggaagcagcagcctg 25500
Db 25441 gctcagcacagacctgtgagacaggtggttgcctctgtgagtggaagcagcagcctg 25500
QY 25501 accttggccttcoggaaatgaaggggcagagcggagccagggcctctttttcaggagct 25560
Db 25501 accttggccttcoggaaatgaaggggcagagcggagccagggcctctttttcaggagct 25560
QY 25561 tgattttgagagcatctggactgctctcccttccctcctcggaggcccttagcagcct 25620
Db 25561 tgattttgagagcatctggactgctctcccttccctcctcggaggcccttagcagcct 25620
QY 25621 gggagcctctgcctcttagagggttccctccatgcattcttttccattttcagctg 25680
Db 25621 gggagcctctgcctcttagagggttccctccatgcattcttttccattttcagctg 25680
QY 25681 tggcctgttggctgtgcgaaggaagggcggttggcgtgctgtgtgagcacatgactgc 25740
Db 25681 tggcctgttggctgtgcgaaggaagggcggttggcgtgctgtgtgagcacatgactgc 25740
QY 25741 atcccttcagctccctgtcccacacccctgcctctgagacatgctctgtctctctatt 25800
Db 25741 atcccttcagctccctgtcccacacccctgcctctgagacatgctctgtctctctatt 25800
QY 25801 gtgtcttctaggtctgttttggagactttccaaatggccaaagctgattcccccactga 25860
Db 25801 gtgtcttctaggtctgtgtttggagactttccaaatggccaaagctgattcccccactga 25860
QY 25861 aagccctgtttgtgcgcacatcgagagatggaggagcgcaggagagagcgggagctgc 25920
Db 25861 aagccctgtttgtgcgcacatcgagagatggaggagcgcaggagagagcgggagctgc 25920
QY 25921 ggcaggtgcggggccctcctctcagggagctggcagcggccttgagagatggggagc 25980
Db 25921 ggcaggtgcggggccctcctctcagggagctggcagcggccttgagagatggggagc 25980
QY 25981 ctcagcagaagcggggccacacagagagcgcacagggccaaagaggtcagagccagtgaa 26040
Db 25981 ctcagcagaagcggggccacacagagagcgcacagggccaaagaggtcagagccagtgaa 26040
QY 26041 gatctgggagacctgaactcagaaggtgtgtcttctgcctccacgcagcaccogta 26100
Db 26041 gatctgggagacctgaactcagaaggtgtgtcttctgcctccacgcagcaccogta 26100
QY 26101 tctgcctccttctgtgtagaagctgaagacagcgttccccccagggagcagctcagata 26160
Db 26101 tctgcctccttctgtgtagaagctgaagacagcgttccccccagggagcagctcagata 26160

XX DE Human ORFX ORF2000 polynucleotide sequence SEQ ID NO:3999.

XX KW Human; open reading frame; ORFX: detection; cytotstatic; hepatotropic;

XX KW vulnary; antipsoariatic; antiparkinsonian; nootropic; neuroprotective;

XX KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;

XX KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

XX KW hypotensive; dermatological; immunosuppressive; antiinflammatory;

XX KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;

XX KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;

XX KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

XX KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

XX KW cholesterol ester storage; systemic lupus erythematosus; infection;

XX KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

XX KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

XX KW bone damage; cartilage damage; antiinflammatory disease; coagulation;

XX KW thrombosis; contraceptive; ss.

OS Homo sapiens.

XX WO200058473-A2.

PN 05-OCT-2000.

PD 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

PA Shimkets RA, Leach M;

PI WPI; 2000-602362/57.

XX P-PSDB; AAB42236.

DR Novel nucleic acids and peptides derived from open reading frame X,

PT useful for treating e.g. cancers, proliferative disorders,

PT neurodegenerative disorders and cardiovascular disease -

XX Claim 5; Page 3179-3180; 5507pp; English.

XX AAC74446 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytotstatic; hepatotropic; vulnary;

CC antipsoariatic; antiparkinsonian; nootropic; neuroprotective;

CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;

CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;

CC antidiabetic; hypotensive; dermatological; immunosuppressive;

CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;

CC antithyroid; and antinaemic. The sequences can be used for determining

CC the presence of or predisposition to, or preventing or treating

CC pathological conditions associated with an ORFX-associated disorder. The

CC nucleic acids can be used to express ORFX proteins in gene therapy

CC vectors. The proteins and nucleic acids may be used to treat cancers,

CC proliferative disorders, neurodegenerative disorders, osteoarthritis,

CC graft vs host disease, cardiovascular disease, diabetes mellitus,

CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus

CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,

CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,

CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,

CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance

CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX SQ Sequence 2546 BP; 652 A; 643 C; 686 G; 564 T; 1 other;

Query Match 1.98; Score 514; DB 21; Length 2546;

Best Local Similarity 99.78; Pred. No. 3.2e-158;

Matches 614; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 25851 cccccactgaaagccctgtttgtcgtggacatcgagagatggaggagcgagggagaag 25910

Db 1892 cccccactgaaagccctgtttgtcgtggacatcgagagatggaggagcgagggagaag 1951

QY 25911 cgggagctggcgaggtgcggcgccctcctgtccaggagagctggcgagggcgctggag 25970

Db 1952 cgggagctggcgaggtgcggcgccctcctgtccaggagagctggcgagggcgctggag 2011

QY 25971 gatgggagctcagcagaagcgccacacagagagccacagagccacagagaggtcaga 26030

Db 2012 gatgggagctcagcagaagcgccacacagagagccacagagagccacagagaggtcaga 2071

QY 26031 gccagtgagatctggagacccctgaactcagaagctgtgtctctctcccccacgca 26090

Db 2072 gccagtgagatctggagacccctgaactcagaagctgtgtctctctcccccacgca 2131

QY 26091 ggcacccgtatctccctcctctgtgtgtagaagctgaagagcacgggtccccagaggca 26150

Db 2132 ggcacccgtatctccctcctctgtgtgtagaagctgaagagcacgggtccccagaggca 2191

QY 26151 gctcagataggtgtgtagagctgtccgaggttggtgtccacataaagcactagtct 26210

Db 2192 gctcagataggtgtgtagagctgtccgaggttggtgtccacataaagcactagtct 2251

QY 26211 atagatgcctcttaggactgtgctggtgcacagcgcggtggagggtgccacacgga 26270

Db 2252 atagatgcctcttaggactgtgctggtgcacagcgcggtggagggtgccacacgga 2311

QY 26271 agcaagcagatgaactaatcttcatttcaaggcagtttttaaaagagctttggaacagac 26330

Db 2312 agcaagcagatgaactaatcttcatttcaaggcagtttttaaaagagctttggaacagac 2371

QY 26331 ggcgcacaccttctctaaatccagcaaatgattccctgcacacacagacaagcagagt 26390

Db 2372 ggcgcacaccttctctaaatccagcaaatgattccctgcacacacagacaagcagagt 2431

QY 26391 aacagatcagtggtgtctaagtgctcagagacttaacgaaaaatatttcagctgcaata 26450

Db 2432 aacagatcagtggtgtctaagtgctcagagacttaacgaaaaatatttcagctgcaata 2491

QY 26451 aagattgagttgcaa 26466

Db 2492 aagattgagttgcaa 2507

RESULT 4

AA52810

ID AA52810 standard; cDNA; 2478 BP.

XX

AC AA52810;

XX

DT 20-SEP-2000 (first entry)

XX

DE Human sulphatase G cDNA.

XX

KW Human sulphatase G; hSG; chromosome 17p11.2; gene therapy; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 1..2478

FT /*tag= a

FT /partial

FT /product= "hSG"

XX

PN WO200034327-A1.

PD 15-JUN-2000.

XX

PF 09-DEC-1999; 99WO-AU01092.

XX

PR 09-DEC-1998; 98AU-0007624.

XX


```
XX PD 06-SEP-2000.
XX XX
XX PF 21-FEB-2000; 2000EP-0200610.
XX XX
XX PR 26-FEB-1999; 99US-0122487.
XX XX
XX PA (GEST ) GENSET.
XX XX
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX XX
XX DR WPI; 2000-500381/45.
XX XX
XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX XX
XX PS Claim 1; SEQ ID 31348; 71pp + CD-ROM; English.
XX XX
XX CC The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX CC identified within the present sequence. The 5' ESTs were prepared from
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences
XX CC derived from the 5' ends of mRNAs and even in those cases where longer
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX CC They are used to obtain upstream regulatory sequences and to design
XX CC expression and secretion vectors.
XX XX
XX SQ Sequence 183 BP; 36 A; 63 C; 38 G; 46 T; 0 other;

Query Match 0.5%; Score 143; DB 21; Length 183;
Best Local Similarity 100.0%; Pred. No. 2.2e-37;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17414 ctacgtcactgcaactcgcctccgcgggttcacgcatctctcgtccacgctcccg 17473
Db 1 ctacgtcactgcaactcgcctccgcgggttcacgcatctctcgtccacgctcccg 60

QY 17474 aatagctggaattacagcgctgctaccacgcccgcgctaatctttgtattttacag 17533
Db 61 aatagctggaattacagcgctgctaccacgcccgcgctaatctttgtattttacag 120

QY 17534 agacgggtttcaccatgttgcca 17556
Db 121 agacgggtttcaccatgttgcca 143

RESULT 7
AAA96365/c
ID AAA96365 standard; DNA: 50000 BP.
XX XX
XX AC AAA96365;
XX XX
XX DT 08-FEB-2001 (first entry)
XX XX
XX DE Polymorphic repeat microsatellite sequences present in the CTLA4 locus.
XX XX
XX KW Autoimmune disease; polymorphic microsatellite repeat; PMR; CD28 gene;
XX KW ICOS gene; CTLA4 gene; costimulatory receptor gene locus; CGRL; lupus;
XX KW insulin-dependent diabetes mellitus; IDDM; Addison's disease; leprosy;
XX KW Graves disease; autoimmune hypothyroidism; myasthenia gravis; thymoma;
XX KW thyroiditis; postpartum thyroiditis; rheumatoid arthritis;
XX KW Hashimoto's disease; coeliac disease; ss.
XX XX
XX OS Homo sapiens.
XX XX
```

```
1
FH FT Location/Qualifiers
FT FT 10141..10177
FT FT /*tag= a
FT FT /note= "sara39/40 microsatellite repeat"
FT FT 11459..11520
FT FT /*tag= b
FT FT /note= "sara33/34 microsatellite repeat"
FT FT 12329..12419
FT FT /*tag= c
FT FT /note= "sara35/36 microsatellite repeat"
FT FT 13527..13567
FT FT /*tag= d
FT FT /note= "sara37/38 microsatellite repeat"
FT FT 24050..24075
FT FT /*tag= e
FT FT /note= "saral1/12 microsatellite repeat"
FT FT 26009..26056
FT FT /*tag= f
FT FT /note= "saral3/14 microsatellite repeat"
FT FT 27317..27350
FT FT /*tag= g
FT FT /note= "saral5/16 microsatellite repeat"
FT FT 30069..30101
FT FT /*tag= h
FT FT /note= "sara21/22 microsatellite repeat"
FT FT 30535..30574
FT FT /*tag= i
FT FT /note= "sara23/24 microsatellite repeat"
FT FT 33714..33758
FT FT /*tag= j
FT FT /note= "sara9/10 microsatellite repeat"
FT FT 43819..43925
FT FT /*tag= k
FT FT /note= "sara31/32 microsatellite repeat"
FT FT 46547..46572
FT FT /*tag= l
FT FT /note= "sara5/6 microsatellite repeat"
FT FT 46828..46875
FT FT /*tag= m
FT FT /note= "sara7/8 microsatellite repeat"
XX XX
XX PN W0200056856-A2.
XX XX
XX PD 28-SEP-2000.
XX XX
XX PF 24-MAR-2000; 2000WO-US07938.
XX XX
XX PR 25-MAR-1999; 99US-0126215.
XX XX
XX PA (GEMY ) GENETICS INST INC.
XX XX
XX PI Ling V, Wu P, Gray GS;
XX XX
XX DR WPI; 2000-628257/60.
XX XX
XX PT Determining predisposition of humans to develop autoimmune disease
XX PT involves detecting polymorphic microsatellite repeat sequence within
XX PT human costimulatory receptor gene locus -
XX XX
XX PS Claim 2; Page 98-113; 160pp; English.
XX XX
XX CC Two human bacterial artificial chromosome (BAC) clones that included
XX CC and flanked the human CTLA-4 locus were cloned and sequenced. The
XX CC sequence data was assembled into a contiguous sequence that is presented
XX CC in AAA96363-68. AAA96363-64 comprise BAC clone 22700, and AAA96365-68
XX CC comprise BAC clone 22608. The sequences contain polymorphic
XX CC microsatellite repeat (PMR) sequences. The specification describes a
XX CC method for determining the predisposition of a human subject to develop
XX CC autoimmune disease. The method comprises detecting a PMR sequence in the
XX CC CD28, ICOS gene or CTLA4 gene of the human costimulatory receptor gene
XX CC locus (HGRL). PMR sequences vary in length among individuals and can be
XX CC amplified to generate products that differ in size. These products can
XX CC then be detected by rapid and convenient high resolution processes. The
```


CC method is useful for determining the predisposition of insulin-dependent
CC diabetes mellitus (IDDM), Addison's disease, Graves disease, autoimmune
CC hypothyroidism, myasthenia gravis, thymoma, lupus, thyroiditis,
CC postpartum thyroiditis, rheumatoid arthritis, Hashimoto's disease,
CC coeliac disease and leprosy. PMR sequences within hCRGL are useful as
CC markers in a variety of assays and in the field of forensic medicine,
CC disease diagnosis and human genome mapping.

XX
SQ Sequence 50000 BP; 14509 A; 9641 C; 10130 G; 15720 T; 0 other;

Query Match 0.3%; Score 84; DB 21; Length 50000;
Best Local Similarity 100.0%; Pred. No. 1.1e-18;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4719 totatttttagtagacgggtttccaccatgttgccaggaatagctcgatctcttgac 4778
|||||

Db 48846 TGTATTTTGTAGACAGCGGGTTTCACCATGTTGCCAGGATAGTCTCGATCTCTTGAC 48787
|||||

Qy 4779 cttgtgatctgccgcctcagcct 4802
|||||

Db 48786 CTTGTGATCTGCCGCTCAGCCT 48763
|||||

RESULT 8

ID AAZ60888 standard; DNA; 17131 BP.

XX
AC AAZ60888;

XX
DT 16-MAY-2000 (first entry)

XX
DE DNA encoding a human geranylgeranyl pyrophosphate synthetase (hGGPPS).

XX
KW Human; geranylgeranyl pyrophosphate synthetase; hGGPPS; chromosome 1;

KW lq42-lq43 locus; prostate cancer; hGGPPS; biallelic marker;

KW mevalonic biosynthetic pathway; ss.

XX
OS Homo sapiens.

XX
FH Key Location/Qualifiers

FT exon 486..546

FT /*tag= a

FT /number= 1

FT intron 547..7291

FT /*tag= b

FT /number= 1

FT exon 633..826

FT /*tag= c

FT /number= 1bis

FT intron 827..7191

FT /*tag= d

FT /number= 1bis

FT exon 7292..7384

FT /*tag= e

FT /number= 2

FT intron 7385..13759

FT /*tag= f

FT /number= 2

FT exon 13760..13830

FT /*tag= g

FT /number= 3

FT intron 13831..14062

FT /*tag= h

FT /number= 3

FT exon 14063..15251

FT /*tag= i

FT /number= 4

XX
PN WO200005382-A2.

XX
PD 03-FEB-2000.

XX

PF 23-JUL-1999; 99WO-TB01353.

XX
PR 23-JUL-1998; 98US-0093940.

XX
PA (GEST) GENSET.

XX
PI Bougueleret L;

XX
DR WPI: 2000-182704/16.

XX
DR P-PSDB; AAY68909.

XX
PT New isolated human geranyl-geranyl pyrophosphate synthetase nucleic
XX acids, used to develop agents for the diagnosis of, e.g. pathologies
XX related to a defect in the mevalonic biosynthetic pathway -

XX
PS Claim 1; Page 72-79; 88pp; English.

XX
CC The present sequence represents a genomic sequence of human
CC geranylgeranyl pyrophosphate synthetase (hGGPPS). The sequence
CC comprises the 5' regulatory region, the exons and introns, and
CC 3' regulatory region. Two differently spliced mRNAs exist for this
CC gene. The first spliced mRNA is derived from a cDNA (AAZ60888) which
CC comprises exons 1, 2, 3 and 4. The second mRNA is derived from a
CC cDNA (AAZ60889) which comprises 1bis, 2, 3, and 4. The hGGPPS gene is
CC located on chromosome 1, at the lq42-lq43 locus. This chromosome 1
CC locus has been shown to carry a predisposing gene for prostate cancer.
CC The nucleic acids encoding hGGPPS can be used for screening for agents
CC which modulate the expression of the hGGPPS gene. Such agents can be
CC used in therapeutic applications. The biallelic markers associated with
CC the hGGPPS gene can be used for the diagnosis of diseases related to
CC an alteration in the regulatory or coding regions of hGGPPS, such as
CC pathologies related to a defect in the mevalonic biosynthetic pathway.
CC The products can also be used for detection, diagnosis and drug
CC screening.

XX
SQ Sequence 17131 BP; 5110 A; 3434 C; 3759 G; 4816 T; 12 other;

Query Match 0.3%; Score 76; DB 21; Length 17131;

Best Local Similarity 100.0%; Pred. No. 5.4e-16;

Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17358 ttttttgagacggaggtttcactctgttgcacaggtgagtgcaatggcgtgatctca 17417
|||||

Db 16721 ttttttgagacggaggtttcactctgttgcacaggtgagtgcaatggcgtgatctca 16780
|||||

Qy 17418 gctcactgcaacctcc 17433
|||||

Db 16781 gctcactgcaacctcc 16796
|||||

RESULT 9

AAZ80231

ID AAZ80231 standard; cDNA; 238 BP.

XX
AC AAZ80231;

XX
DT 07-APR-2000 (first entry)

XX
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:315.

XX
DE Human; gene expression product; diagnosis; tumour; colon cancer;

XX
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;

XX
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;

XX
OS Homo sapiens.

XX
PN WO9964576-A2.

XX
PD 16-DEC-1999.

XX
PF 09-JUN-1999; 99WO-TB01062.

XX 10-JUN-1998; 98US-0088801.
XX (FARB) BAYER CORP.
XX
XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
XX WPI; 2000-087220/07.
XX
XX Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer -
XX
XX Claim 15; Page 258; 469pp; English.
XX
XX AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g.
CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC can be used for developing agents for the diagnosis and treatment of
CC disorders involving unwanted cell proliferation, such as neoplasia,
CC dysplasia or hyperplasia.
XX
SQ Sequence 238 BP; 55 A; 57 C; 69 G; 57 T; 0 other;

Query Match 0.3%; Score 72; DB 21; Length 238;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1085 acctgcaggtggtggcagcggttagccgggactcggcgccgcgcgtctctctccg 1144
Db 1 acctgcaggtggtggcagcggttagccgggactcggcgccgcgcgtctctctccg 60
QY 1145 agttcaaccggt 1156
Db 61 agttcaaccggt 72

RESULT 10
AAC02520
ID AAC02520 standard; cDNA; 215 BP.
XX
XX AAC02520;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 2518.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GENT) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX WPI; 2000-500381/45.
DR P-PSDB; AAG02514.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 2518; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
XX vectors.
XX
SQ Sequence 215 BP; 45 A; 47 C; 80 G; 43 T; 0 other;

Query Match 0.3%; Score 71; DB 21; Length 215;
Best Local Similarity 100.0%; Pred. No. 5.4e-14;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8353 agaagagatagcagatcttcagaaaagctgatggaagccgggtgcagtggtcacgcct 8412
Db 1 agaagagatagcagatcttcagaaaagctgatggaagccgggtgcagtggtcacgcct 60
QY 8413 gtaatccacgc 8423
Db 61 gtaatccacgc 71

RESULT 11
AAZ94762/c
ID AAZ94762 standard; cDNA; 2923 BP.
XX
XX AAZ94762;
XX
XX 01-AUG-2000 (first entry)
XX
XX Human ATP binding cassette ABCG1 (ABC8) 5' untranslated region.
XX
XX ABCG1; ABC8; ATP binding cassette; human; phagocyte; cholesterol;
KW phospholipid; transporter; inflammation; atherosclerosis;
KW lipid disorder; dyslipidemia; psoriasis; lupus erythematosus;
KW diagnosis; gene therapy; chromosome 21q22.3; ss.
XX
XX Homo sapiens.
XX
XX WO200018912-A2.
XX
XX 06-APR-2000.
XX
XX 21-SEP-1999; 99WO-EP06991.
XX
XX 25-SEP-1998; 98US-0101706.
XX
XX (FARB) BAYER AG.
XX
XX Schmitz G, Klucken J;
PI
XX WPI; 2000-293151/25.
XX
XX Adenosine triphosphate binding proteins useful for identifying agents
PT for treating atherosclerosis and other inflammatory disorders -
XX
XX Claim 9; Page 143-144; 154pp; English.

XX The present sequence is that of the 5' untranslated region of the
CC human ATP binding cassette protein ABCG1 (ABC8) gene. ABCG1 is
CC a cholesterol switch. It is up-regulated by macrophage colony
CC stimulating factor dependent phagocytic differentiation, and
CC expression is massively induced by cholesterol loading and almost
CC completely set back to differentiation dependent levels by HDL3.
CC ABCG1 is the human homologue of the Drosophila white gene.
CC Sequencing of the promoter revealed important transcription factor
CC binding sites relevant for phagocytic differentiation and lipid
CC sensitivity. ABCG1 was also identified as a cholesterol
CC transporter and modulator of choline-containing phospholipids
CC (phosphatidylcholine, sphingomyelin). The invention provides
CC cholesterol-sensitive ABC genes (see AA294734-63) that can be used
CC for diagnostic and therapeutic applications, and for biochemical or
CC cell-based assays to screen for pharmacologically active compounds
CC useful for the treatment of lipid disorders, atherosclerosis or
CC other inflammatory diseases such as psoriasis and lupus
CC erythematosus.
XX
SQ Sequence 2923 BP; 593 A; 873 C; 803 G; 641 T; 13 other;

Query Match 0.2%; Score 60; DB 21; Length 2923;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18972 ggagatcgagaccatcctgctaacaatgatgaaccccgctctactataacatacaaaa 19031
Db 1569 GGAGATCGAGACCCTCTGGCTACATGATGAACCCCGTCTCTACTATAAATAACAAAA 1510

RESULT 12
AAZ94784/C
ID AAZ94784 standard; CDNA; 2923 BP.
AC AAZ94784;
XX
DT 01-AUG-2000 (first entry)
XX
DE Human ATP binding cassette ABCG1 (ABC8) 5' untranslated region.
XX
KW ABCG1; ABC8; ATP binding cassette; human; phagocyte; cholesterol;
KW phospholipid; transporter; inflammation; atherosclerosis;
KW lipid disorder; dyslipidemia; psoriasis; lupus erythematosus;
KW diagnosis; gene therapy; chromosome 21q22.3; ss.
XX
OS Homo sapiens.
XX
PN WO200018912-A2.
XX
PD 06-APR-2000.
XX
PF 21-SEP-1999; 99WO-EP06991.
XX
PR 25-SEP-1998; 98US-0101706.
XX
PA (FARB) BAYER AG.
XX
PI Schmitz G, Klucken J;
XX
DR WPI; 2000-293151/25.
XX
PT Adenosine triphosphate binding proteins useful for identifying agents
PT for treating atherosclerosis and other inflammatory disorders -
XX
PS Claim 9; Page 151-152; 154pp; English.
XX
CC The present sequence is that of the 5' untranslated region and exon
CC 1 of the human ATP binding cassette protein ABCG1 (ABC8) gene.
CC ABCG1 is a cholesterol switch. It is up-regulated by macrophage
CC colony stimulating factor dependent phagocytic differentiation, and
CC expression is massively induced by cholesterol loading and almost

CC completely set back to differentiation dependent levels by HDL3.
CC ABCG1 is the human homologue of the Drosophila white gene.
CC Sequencing of the promoter revealed important transcription factor
CC binding sites relevant for phagocytic differentiation and lipid
CC sensitivity. ABCG1 was also identified as a cholesterol
CC transporter and modulator of choline-containing phospholipids
CC (phosphatidylcholine, sphingomyelin). The invention provides
CC cholesterol-sensitive ABC genes (see AA294734-63) that can be used
CC for diagnostic and therapeutic applications, and for biochemical or
CC cell-based assays to screen for pharmacologically active compounds
CC useful for the treatment of lipid disorders, atherosclerosis or
CC other inflammatory diseases such as psoriasis and lupus
CC erythematosus.
XX
SQ Sequence 2923 BP; 593 A; 873 C; 803 G; 641 T; 13 other;

Query Match 0.2%; Score 60; DB 21; Length 2923;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18972 ggagatcgagaccatcctgctaacaatgatgaaccccgctctactataacatacaaaa 19031
Db 1569 GGAGATCGAGACCCTCTGGCTACATGATGAACCCCGTCTCTACTATAAATAACAAAA 1510

RESULT 13
AAZ75924/C
ID AAZ75924 standard; DNA; 9721 BP.
XX
AC AAZ75924;
XX
DT 29-JUL-1999 (first entry)
XX
DE Human interleukin 1B gene.
XX
KW Human; interleukin 1; IL-1B; IL-1A; IL-1RN; diagnosis; detection;
KW chronic obstructive airway disease; chronic bronchitis; emphysema;
KW asthma; chronic bronchiolitis; proinflammatory haplotype; ss.
XX
OS Homo sapiens.
XX
PN WO9924615-A2.
XX
PD 20-MAY-1999.
XX
PF 09-NOV-1998; 98WO-US23721.
XX
PR 12-JAN-1998; 98US-0005923.
PR 07-NOV-1997; 97GB-0023553.
XX
PA (MEDI-) MEDICAL SCI SYSTEMS INC.
XX
PI Barnes PJ, Duff GW, Giovine M, Lim S;
XX
DR WPI; 1999-327420/27.
XX
PT Genotyping nucleic acid samples for interleukin-1 (IL-1)
PT proinflammatory haplotype alleles, useful for predicting
PT susceptibility to developing chronic obstructive airway disease
XX
PS Example 1; Fig 2; 37pp; English.
XX
CC The present invention describes genotyping a nucleic acid sample from a
CC subject to determine at least one allele of an interleukin-1 (IL-1)
CC proinflammatory haplotype. A method has also been described for
CC determining a subject's susceptibility to developing chronic obstructive
CC airway disease (COAD) or for predicting the rapidity or ultimate
CC progression of a COAD in the subject by: (a) obtaining a nucleic acid
CC sample from the subject; and (b) detecting at least one allele of an
CC IL-1 proinflammatory haplotype in the sample, where detection of at
CC least one of these alleles indicates that the patient has an increased
CC susceptibility to developing COAD. The method is useful for determining

CC the susceptibility of subjects to developing chronic obstructive airway
CC disease or for predicting the rapidity or ultimate progression of
CC chronic obstructive airway disease (COAD). COAD can be asthma, emphysema,
CC chronic bronchitis or chronic bronchiolitis. The method provides for
CC early identification of chronic obstructive airway disease (COAD),
CC facilitating administration of appropriate treatment at the earliest
CC stage, thereby increasing the probability of a positive outcome. The
CC present sequence represents the human IL-1B gene.

XX Sequence 9721 BP; 2662 A; 2328 C; 2121 G; 2608 T; 2 other;

SQ Query Match 0.2%; Score 60; DB 20; Length 9721;

Best Local Similarity 100.0%; Pred. No. 9.5e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11740 ttttttgagacagagtctactctgtcgccaggctggagtgagtgagcagcagctctgg 11799
|||||
Db 956 TTTTGTGAGACAGAGTCTCACTCTGTGCGCCAGGCTGGAGTGCGACGATCTCGG 897

RESULT 14

AAF20945/c

ID AAF20945 standard; DNA; 9721 BP.

XX AC AAF20945;

DT 14-MAR-2001 (first entry)

XX Human interleukin-1 polynucleotide fragment #2512.

DE Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.

XX Homo sapiens.

XX WO200062736-A2.

XX 26-OCT-2000.

XX 24-MAR-2000; 2000WO-US08020.

XX 06-APR-1999; 99US-0127958.

XX (UYEC-) UNIV EAST CAROLINA.

XX (NYCE/) NYCE J W.

XX Nyce JW;

XX WPI; 2000-679539/66.

XX Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -

PS Disclosure; Page 230-232; 1592pp; English.

XX The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and/or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and

CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system peptide
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.

SQ Sequence 9721 BP; 2661 A; 2328 C; 2122 G; 2608 T; 2 other;

Query Match 0.2%; Score 60; DB 21; Length 9721;

Best Local Similarity 100.0%; Pred. No. 9.5e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11740 ttttttgagacagagtctactctgtcgccaggctggagtgagtgagcagcagctctgg 11799
|||||
Db 956 TTTTGTGAGACAGAGTCTCACTCTGTGCGCCAGGCTGGAGTGCGACGATCTCGG 897

RESULT 15

AAC63768/c

ID AAC63768 standard; DNA; 9721 BP.

XX AC AAC63768;

XX 08-FEB-2001 (first entry)

XX Human IL-1B gene.

XX Human; IL-1B; interleukin-1B; cytostatic; antiinflammatory;

XX immunosuppressive; dermatological; antimicrobial; antiarthritic;

XX IL-1 receptor antagonist; tumour necrosis factor alpha antagonist

XX interstitial lung disease; interstitial pneumonia; pulmonary fibrosis;

XX rheumatoid arthritis; systemic lupus erythematosus; Sjogren's syndrome;

XX systemic sclerosis; dermatomyocitis; chromosome 2; ds.

XX Homo sapiens.

XX WO200060117-A2.

XX 12-OCT-2000.

XX 31-MAR-2000; 2000WO-US08492.

XX 02-APR-1999; 99US-0286108.

XX (INTE-) INTERLEUKIN GENETICS INC.

XX Duff GW, Di Giovine FS, Whyte M;

XX WPI; 2000-656234/63.

XX Method for predicting the risk of interstitial lung disease, comprising
PT detecting an interleukin-1 receptor antagonist allele and tumor
PT necrosis alpha allele or an allele in linkage disequilibrium with
PT either of these alleles -

XX Claim 6; Fig 2; 102pp; English.

XX The present sequence is provided in a specification relating to a method
CC for determining whether a subject has or is predisposed to develop an
CC interstitial lung disease. The method involves detecting an interleukin-1
CC receptor antagonist (IL-1RN) (+2018) allele 2, a tumour necrosis alpha
CC (TNF-A)(-308) allele 2, or an allele in linkage disequilibrium with
CC either of these two alleles. The method may be used to determine whether
CC a subject has or is predisposed to develop an interstitial pneumonia or a
CC pulmonary fibrosis and other disorders such as rheumatoid arthritis,
CC systemic lupus erythematosus, Sjogren's syndrome, systemic sclerosis,
CC dermatomyositis. The method is also used for identifying molecules
CC which can be used as therapeutics for treating interstitial lung disease.
XX
SQ Sequence 9721 BP; 2661 A; 2328 C; 2122 G; 2608 T; 2 other;

Query Match 0.2%; Score 60; DB 21; Length 9721;
Best Local Similarity 100.0%; Pred. No. 9.5e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11740 ttttttgacagagtcctcactctgtcccgagctggagtgagtgagtcacgacatctcgg 11799
|||||
Db 956 TTTTGTGACAGAGTCTCCTCTGTGCGCCAGGCTGGAGTGCAGTGCAGATCTCGG 897

RESULT 16
AAA50174/c
ID AAA50174 standard; DNA; 9721 BP.
XX
AC AAA50174;
XX
DT 07-NOV-2000 (first entry)
XX
DE Human interleukin-1 beta allele 1 (+6912).
XX
KW Interleukin-1 beta; IL-1b; human; polymorphism; inflammation;
KW coronary artery disease; osteoporosis; nephropathy;
KW alopecia areata; Graves disease; systemic lupus erythematosus;
KW lichen sclerosis; ulcerative colitis; diabetic retinopathy;
KW periodontal disease; juvenile chronic arthritis; psoriasis;
KW insulin dependent diabetes; asthma; lung fibrosis;
KW chronic inflammatory liver disease; rheumatoid arthritis;
KW chronic inflammatory lung disease; antiinflammatory; osteopathic;
KW dermatological; immunosuppressive; antidiabetic; antihypertensive;
KW antiarthritic; antirheumatic; antiasthmatic; antipsoriatic;
KW hepatotropic; antiulcer; diagnosis; therapy; ds.
XX
OS Homo sapiens.

XX
XX Key Location/Qualifiers
FT variation replace(8904,G)
FT /*tag= a
FT /note= "IL-1B allele 1 (+6912)"
XX
XX W0200047619-A1.
XX
XX 17-AUG-2000.
XX
XX 10-FEB-2000; 2000W0-US03443.
XX
XX 10-FEB-1999; 99US-0247874.
XX
XX (INTE-) INTERLEUKIN GENETICS INC.
XX
XX Duff GW, Di Giovine FS;
XX WPI; 2000-558192/51.
XX
XX Novel methods and nucleic acids for diagnosing and treating disorders
XX associated with high levels of interleukin 1beta, especially
XX inflammatory diseases -
XX
XX Disclosure; Fig 1; 74pp; English.

XX
CC The present sequence is that of human interleukin-1 beta (IL-1B)
CC allele 1 (+6912), which is a form of the IL-1B gene that contains
CC cytosine at position +6912; IL-1B allele 2 (+6912) has guanine at
CC this position (see AAA50175). The invention is based on the
CC identification of this novel allele at marker +6912 of the IL-1B
CC gene. The C to G transition occurs within the 3' untranslated
CC region of the IL-1B gene and results in an increased level of IL-1B
CC protein. Individuals homozygous for the IL-1B allele 2 (+6912)
CC accumulate approximately 4 times more immunoreactive IL-1B protein
CC than homozygotes for IL-1B allele 1 (+6912). Methods and kits are
CC provided for detecting IL-1B allele 2 (+6912), or an allele in
CC linkage disequilibrium with an IL-1B allele 2 (+6912), and
CC thereby determining a patient's susceptibility to developing
CC inflammatory disorders, especially coronary artery disease,
CC osteoporosis, nephropathy in diabetes mellitus, alopecia areata,
CC Graves disease, systemic lupus erythematosus, lichen sclerosis,
CC ulcerative colitis, diabetic retinopathy, periodontal disease,
CC juvenile chronic arthritis, psoriasis, insulin dependent diabetes,
CC asthma, chronic inflammatory liver disease, chronic inflammatory
CC lung disease, lung fibrosis, and rheumatoid arthritis (claimed).
CC Identification of the IL-1B allele 2 (+6912) and its involvement in
CC IL-1B overproduction also enables screening assays for identifying
CC IL-1B antagonists that can be used to treat conditions associated
CC with IL-1B allele 2 (+6912). Transgenic animals are also claimed,
CC and can be used to identify IL-1B agonists and antagonists, or
CC to confirm the safety and efficacy of candidate therapeutics.
XX
SQ Sequence 9721 BP; 2661 A; 2328 C; 2122 G; 2608 T; 2 other;

Query Match 0.2%; Score 60; DB 21; Length 9721;
Best Local Similarity 100.0%; Pred. No. 9.5e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11740 ttttttgacagagtcctcactctgtcccgagctggagtgagtcacgacatctcgg 11799
|||||
Db 956 TTTTGTGACAGAGTCTCCTCTGTGCGCCAGGCTGGAGTGCAGTGCAGATCTCGG 897

RESULT 17
AAA50175/c
ID AAA50175 standard; DNA; 9721 BP.
XX
AC AAA50175;
XX
DT 07-NOV-2000 (first entry)
XX
DE Human interleukin-1 beta allele 2 (+6912).
XX
KW Interleukin-1 beta; IL-1B; human; polymorphism; inflammation;
KW coronary artery disease; osteoporosis; nephropathy;
KW alopecia areata; Graves disease; systemic lupus erythematosus;
KW lichen sclerosis; ulcerative colitis; diabetic retinopathy;
KW periodontal disease; juvenile chronic arthritis; psoriasis;
KW insulin dependent diabetes; asthma; lung fibrosis;
KW chronic inflammatory liver disease; rheumatoid arthritis;
KW chronic inflammatory lung disease; antiinflammatory; osteopathic;
KW dermatological; immunosuppressive; antidiabetic; antihypertensive;
KW antiarthritic; antirheumatic; antiasthmatic; antipsoriatic;
KW hepatotropic; antiulcer; diagnosis; therapy; ds.
XX
OS Homo sapiens.

XX
XX Key Location/Qualifiers
FT variation replace(8904,C)
FT /*tag= a
FT /note= "IL-1B allele 2 (+6912)"
XX
XX W0200047619-A1.
XX
XX 17-AUG-2000.
XX

PF 10-FEB-2000; 2000WO-US03443.
XX
PR 10-FEB-1999; 99US-0247874.
XX
PA (INTE-) INTERLEUKIN GENETICS INC.
XX
XX Duff GW, Di Giovine FS;
PI WPI; 2000-558192/51.
XX
DR Novel methods and nucleic acids for diagnosing and treating disorders
XX associated with high levels of interleukin ibeta, especially
PT inflammatory diseases -
XX
XX Claim 34; Fig 2; 74pp; English.
XX
XX The present sequence is that of human interleukin-1 beta (IL-1B)
CC allele 2 (+6912), which is a form of the IL-1B gene that contains
CC guanine at position +6912; IL-1B allele 1 (+6912) has cytosine at
CC this position (see AAA50174). The invention is based on the
CC identification of this novel allele at marker +6912 of the IL-1B
CC gene. The C to G transition occurs within the 3' untranslated
CC region of the IL-1B gene and results in an increased level of IL-1B
CC protein. Individuals homozygous for the IL-1B allele 2 (+6912)
CC accumulate approximately 4 times more immunoreactive IL-1B protein
CC than homozygotes for IL-1B allele 1 (+6912). Methods and kits are
CC provided for detecting IL-1B allele 2 (+6912), or an allele in
CC linkage disequilibrium with an IL-1B allele 2 (+6912), and
CC thereby determining a patient's susceptibility to developing
CC inflammatory disorders, especially coronary artery disease,
CC osteoporosis, nephropathy in diabetes mellitus, alopecia areata,
CC Graves disease, systemic lupus erythematosus, lichen sclerosis,
CC ulcerative colitis, diabetic retinopathy, periodontal disease,
CC juvenile chronic arthritis, psoriasis, insulin dependent diabetes,
CC asthma, chronic inflammatory liver disease, chronic inflammatory
CC lung disease, lung fibrosis, and rheumatoid arthritis (claimed).
CC Identification of the IL-1B allele 2 (+6912) and its involvement in
CC IL-1B overproduction also enables screening assays for identifying
CC IL-1B antagonists that can be used to treat conditions associated
CC with IL-1B allele 2 (+6912). Transgenic animals are also claimed,
CC and can be used to identify IL-1B agonists and antagonists, or
CC to confirm the safety and efficacy of candidate therapeutics.
XX
SQ Sequence 9721 BP; 2661 A; 2327 C; 2123 G; 2608 T; 2 other;

Query Match 0.2%; Score 60; DB 21; Length 9721;
Best Local Similarity 100.0%; Pred. No. 9.5e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11740 ttttttgagacagagtctcactctgtccgccagctggagtcagtgagcagcatctcgg 11799
Db 956 TTTTGTGAGACAGAGTCTCACTGTGCGCCAGGCTGGAGTGCAGTGGCAGATCTCGG 897

RESULT 18
AAA34823/C
ID AAA34823 standard; DNA; 9721 BP.
XX
AC AAA34823;
XX
XX 28-JUL-2000 (first entry)
XX
XX Human adenosine receptor related polynucleotide SEQ ID NO:2512.
XX
XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
XX phosphorothioate; impaired respiration; inflammation; allergy;
XX allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
XX antiallergic; antiasthmatic; cytotstatic; analgesic; impaired airway;
XX lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
XX respiratory distress syndrome; pain; cystic fibrosis; emphysema;
XX pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
XX cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX Homo sapiens.
OS
XX WO200009525-A2.
PN
XX 24-FEB-2000.
PD
XX 03-AUG-1999; 99WO-US17712.
XX
PF 03-AUG-1998; 98US-0095212.
PR
XX (UYEC-) UNIV EAST CAROLINA.
PA
XX Nyce JW;
PI
XX WPI; 2000-205971/18.
DR
XX New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 673-675; 1343pp; English.
XX
XX The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antisthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 9721 BP; 2661 A; 2328 C; 2122 G; 2608 T; 2 other;

Query Match 0.2%; Score 60; DB 21; Length 9721;
Best Local Similarity 100.0%; Pred. No. 9.5e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11740 ttttttgagacagagtctcactctgtccgccagctggagtcagtgagcagcatctcgg 11799
Db 956 TTTTGTGAGACAGAGTCTCACTGTGCGCCAGGCTGGAGTGCAGTGGCAGATCTCGG 897

RESULT 19
AAF27666/C
ID AAF27666 standard; DNA; 9721 BP.
XX
AC AAF27666;
XX
XX 02-APR-2001 (first entry)
DT
XX IL-1B DNA.
DE
XX IL-1; interleukin; inflammation; infection; ds.

```
XX OS Unidentified.
XX PN WO200100880-A2.
XX PD 04-JAN-2001.
XX PF 30-JUN-2000; 2000WO-US18318.
XX PR 30-JUN-1999; 99US-0345217.
XX PA (INTE-) INTERLEUKIN GENETICS INC.
XX PI Duff GW, Cox A, Camp NJ, Di Giovine FS;
XX PS WPI; 2001-102903/11.
XX CC Determining whether a subject has or is predisposed to disease
XX CC associated with IL-1 polymorphism involves determining presence of
XX CC marker or allele comprising IL-1 inflammatory haplotype -
XX PS Disclosure; Fig 4; 84pp; English.
XX CC The present invention relates to a new method for determining whether
XX CC a subject has or is predisposed to developing a disease or condition
XX CC that is associated with an IL (interleukin)-1 inflammatory haplotype,
XX CC comprises detecting at least one allele of the haplotype, where the
XX CC presence of the allele indicates that the subject is predisposed to
XX CC the development or has the disease or condition.
XX CC The method is useful for determining whether a subject has or is
XX CC predisposed to inflammatory disease, a degenerative disease, an
XX CC immunological disorder, an infectious disease, trauma induced disease,
XX CC or cancer. The above conditions associated with an IL-1 inflammatory
XX CC haplotype can be treated or prevented by administering a therapeutic
XX CC that compensates for a causative mutation that is in linkage
XX CC disequilibrium with at least one IL-1 polymorphism.
XX SQ Sequence 9721 BP; 2661 A; 2328 C; 2122 G; 2607 T; 3 other;

Query Match 0.2%; Score 60; DB 22; Length 9721;
Best Local Similarity 100.0%; Pred. No. 9.5e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11740 ttttttgagacagagtcactctgtcgccaggctgagtgagtcagtgccacgatctcg 11799
Db 956 TTTTGTGACAGAGTCTCAGTCTGTGCGCCAGGCTGGAGTGCAGTGGCAGCATCTCGG 897

RESULT 20
AAC91434/c
ID AAC91434 standard; DNA; 9721 BP.
XX AC AAC91434;
XX DT 20-MAR-2001 (first entry)
XX DE Human IL-1B nucleotide sequence.
XX KW Human; IL-1A; interleukin-lalpha; IL-1B; interleukin-lbeta; IL-1RN;
XX KW interleukin-1 receptor antagonist; vasotropic; antiinflammatory;
XX KW hypotensive; anticoagulant; antilipemic; arterial restenosis;
XX KW restenosis associated allele; RAA; occlusive cardiovascular disorder;
XX KW restenosis detection; ds.
XX OS Homo sapiens.
XX PN WO200071753-A2.
XX PD 30-NOV-2000.
XX PF 24-MAY-2000; 2000WO-US14299.
XX
```

```
PR 24-MAY-1999; 99US-0317674.
PR 01-NOV-1999; 99US-0431352.
XX PA (INTE-) INTERLEUKIN GENETICS INC.
XX PI Kornman KS, Duff GW, Crossman DC, Francis SE, Stephenson K;
XX PS WPI; 2001-025173/03.
XX PT Diagnosing or determining susceptibility to developing restenosis
XX PT involves detecting restenosis associated allele in a nucleic acid
XX PT sample -
XX PS Disclosure; Fig 2; 129pp; English.
XX CC The present sequence is given in a specification relating to a method for
XX CC determining whether a subject has or is predisposed to developing an
XX CC arterial restenosis. The method comprises detecting a restenosis
XX CC associated allele (RAA) in a nucleic acid sample from the subject, where
XX CC detection of the RAA indicates that the subject has or is predisposed to
XX CC the development of a restenosis. The restenosis associated allelic
XX CC pattern permits the diagnosis of occlusive cardiovascular disorder. The
XX CC diagnosis allows the most suitable treatment methods for restenosis to be
XX CC used e.g. selecting therapies for initial vascular stenosis most likely
XX CC to avoid subsequent stenoses. The detection methods identify restenosis
XX CC therapeutics, agonists and antagonists, (proteins, peptides,
XX CC peptidomimetics, small molecules or nucleic acids, e.g. anti-sense,
XX CC ribozyme and triplex nucleic acids) which are used to treat restenosis.
XX SQ Sequence 9721 BP; 2661 A; 2328 C; 2122 G; 2608 T; 2 other;

Query Match 0.2%; Score 60; DB 22; Length 9721;
Best Local Similarity 100.0%; Pred. No. 9.5e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11740 ttttttgagacagagtcactctgtcgccaggctgagtgagtcagtgccacgatctcg 11799
Db 956 TTTTGTGACAGAGTCTCAGTCTGTGCGCCAGGCTGGAGTGCAGTGGCAGCATCTCGG 897

RESULT 21
AAAX02996
ID AAAX02996 standard; DNA; 10620 BP.
XX AC AAAX02996;
XX DT 22-JUN-1999 (first entry)
XX DE Human IL-1ra BAC contiguous DNA sequence 41.
XX KW Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
XX KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
XX KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
XX KW growth factors; treatment; IL-1 receptor complex; BAC; ss.
XX OS Homo sapiens.
XX PN WO9906426-A1.
XX PD 11-FEB-1999.
XX PF 03-AUG-1998; 98WO-US16102.
XX PR 02-JUL-1998; 98US-0091650.
XX PR 04-AUG-1997; 97US-0054646.
XX PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX PI Pan Y;
XX PD WPI; 1999-153692/13.
XX
```


CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.

SQ Sequence 29433 BP; 8714 A; 6519 C; 5920 G; 8278 T; 2 other;

Query Match 0.2%; Score 60; DB 21; Length 29433;

Best Local Similarity 100.0%; Pred. No. 7.6e-11;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11740 ttttttgacagagtgctcactctgtcgccaggctgagtgacgtgacgacgtctcgg 11799
|||||
DB 14423 TTTTGTGACAGAGTCTCACTCTGCGCCAGGCTGGAGTGCAGTGGCAGCATCTCGG 14364

RESULT 24

AAA34828/c

ID AAA34828 standard; DNA; 29433 BP.

AC AAA34828;

DT 28-JUL-2000 (first entry)

DE Human adenosine receptor related polynucleotide SEQ ID NO:2517.

KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX Homo sapiens.

OS WO200009525-A2.

PN 24-FEB-2000.

XX 03-AUG-1999; 99WO-US17712.

PF 03-AUG-1998; 98US-0095212.

PR (UYEC-) UNIV EAST CAROLINA.

PA Nyce JW;

PI WPI; 2000-205971/18.

DR New antisense oligonucleotides useful for treating e.g. pulmonary
XX vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -

XX

PS Disclosure; Page 677-684; 1343pp; English.

CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC carcinomas, and cancers which may metastasize to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.

XX Sequence 29433 BP; 8714 A; 6519 C; 5920 G; 8278 T; 2 other;

Query Match 0.2%; Score 60; DB 21; Length 29433;

Best Local Similarity 100.0%; Pred. No. 7.6e-11;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11740 ttttttgacagagtgctcactctgtcgccaggctgagtgacgtgacgacgtctcgg 11799
|||||
DB 14423 TTTTGTGACAGAGTCTCACTCTGCGCCAGGCTGGAGTGCAGTGGCAGCATCTCGG 14364

RESULT 25

AAF21437/c

ID AAF21437 standard; DNA; 209273 BP.

XX AAF21437;

DT 14-MAR-2001 (first entry)

DE Human factor-related antisense polynucleotide #3004.

KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.

XX Homo sapiens.

OS WO2000062736-A2.

PN 26-OCT-2000.

XX 24-MAR-2000; 2000WO-US08020.

PR 06-APR-1999; 99US-0127958.

XX (UYEC-) UNIV EAST CAROLINA.

```
PA (NYCE/) NYCE J W.
XX
XX
PI Nyce JW;
XX
XX WPI; 2000-679539/66.
XX
XX Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
PT
XX
XX Disclosure; Page 55-100; 1592pp; English.
XX
XX The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and/or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (ARDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
XX Sequence 209273 BP; 59273 A; 45843 C; 42459 G; 61176 T; 522 other;
SQ

Query Match 0.2%; Score 60; DB 21; Length 209273;
Best Local Similarity 100.0%; Pred. No. 5,2e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11740 ttttttgagacagagtctcaactctgtgccagggtggagtgagtgagtcgacgatctcg 11799
|||||
Db 186208 TTTTGTGACACAGAGTCTCAGTCTGTGCGCCAGGTGGAGTGCAGTGCCAGCATCTCGG 186149

RESULT 26
AAC21264
ID AAC21264 standard; cDNA; 310 BP.
XX
XX AAC21264;
AC
XX 06-OCT-2000 (first entry)
DT
XX Human secreted protein 5' EST, SEQ ID NO: 25339.
DE
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
PN
XX
XX 06-SEP-2000.
PD
XX

Query Match 0.2%; Score 59; DB 21; Length 310;
Best Local Similarity 100.0%; Pred. No. 4e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11122 tggctaattttgtgttttagtagagacggggtttccaccatgttgccaggctggctt 11180
|||||
Db 101 tggctaattttgtgttttagtagagacggggtttccaccatgttgccaggctggctt 159

RESULT 27
AAT48737/C
ID AAT48737 standard; DNA; 5581 BP.
XX
XX AAT48737;
AC
XX 28-MAY-1997 (first entry)
DT
XX Human leucocyte specific transcript LST-1 gene.
DE
XX Leucocyte specific transcript; LST-1; cytokine; immunoregulator;
KW lymphoma; U-937; antitumour; tumour; gene therapy; diagnosis; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH exon 48..162
FT /tag= a
FT /note= "exon 1A"
FT intron 163..543
FT /tag= b
FT exon 544..652
FT /tag= c
FT /note= "exon 1B"
FT intron 653..1043
FT /tag= d
FT exon 1044..1162
FT /tag= e
FT /note= "exon 2"
FT /codon_start= 1144..1146
FT
```


CC immunoregulatory protein (AAW07680), leukocyte specific transcript-1
CC (LST-1), which is characterised in that its prodn. in U937 cell
CC lines is stimulated by interferon-gamma by a factor of over 1000
CC and that it binds to the surface of leukocytes. The gene was obtd.
CC from a human B-cell line CAH cosmid library by hybridization to a
CC TNFA probe. Alternative splicing yields a 97-amino acid LST-1
CC isoform (AAW07681). An LST-1 cDNA clone (AA745452) has also been
CC isolated. LST-1 nucleic acids can be used for recombinant protein
CC prodn., as probes, and in the gene therapy of e.g. tumours, esp.
CC using methods based on homologous recombination.
XX
SQ Sequence 5581 BP; 1410 A; 1441 C; 1542 G; 1188 T; 0 other;

Query Match 0.2%; Score 59; DB 18; Length 5581;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10490 gctaattttgtatttttagtagagacgggtttccaccatgttggccaggatgtct 10548
|||||
Db 4279 GCTAAATTTGTATTTTGTAGTAGACGGGTTTCACCATGTTGCCAGGATGGTCT 4221
|||||

RESULT 29
AAW83427
ID AAX83427 standard; DNA; 21721 BP.
XX
AC AAX83427;
XX
DT 31-AUG-1999 (first entry)
XX
DE Human lipolysis stimulated receptor genomic sequence.
XX
KW Human; lipolysis stimulated receptor; LSR; lipoprotein; free fatty acid;
KW cytokine; probe; primer; amplification; hybridisation; detection; leptin;
KW allele; variant; mutation; deletion; loss of heterozygosity; chylomicron;
KW transgenic animal; gene expression; triglyceride; eating disorder;
KW obesity; atheromatosis; atherosclerosis; hypertension; diabetes;
KW anorexia; metabolism; ds.
XX
OS Homo sapiens.
XX
PN WO9907737-A2.
XX
PD 18-FEB-1999.
XX
PF 06-AUG-1998; 98WO-IB01257.
XX
PR 22-APR-1998; 98FR-0005032.
PR 06-AUG-1997; 97FR-0010088.
XX
PA (GEST) GENSET.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bihaïn B, Bougueleret L, Yen-Potin F;
XX
DR WPI; 1999-190035/16.
XX
PT Lipolysis stimulated receptor involved in leptin metabolism - and
PT controlling its activity for treatment of obesity, anorexia or
PT abnormal cytokine metabolism

XX
PS Claim 40; Page 258-277; 279pp; French.
XX
CC This sequence represents the human lipolysis stimulated receptor (LSR)
CC genomic sequence. The novel LSR binds lipoproteins in presence of free
CC fatty acids (FFA) and binds cytokines in absence of FFA. The nucleic
CC acid and its fragments are used as probes and primers for detection
CC and/or amplification of LSR genes; for production of recombinant LSR; for
CC detecting allelic variants, mutations, deletions, loss of heterozygosity
CC and genetic abnormalities in the gene. LSR, recombinant cells and
CC transgenic animals are used to screen for chemical interacting with LSR,
CC also to study expression and activity of LSR and its interactions. The

CC chemicals, and leptin, are used to modulate the number of LSR in a cell,
CC its recycling rate and/or specificity of receptor activity, particularly
CC for reducing the level of leptin, lipoproteins, chylomicrons and/or
CC triglycerides. The chemicals are thus useful for treating eating
CC disorders, particularly obesity (and related diseases such as
CC atheromatosis, atherosclerosis, hypertension and diabetes) or anorexia,
CC also disease associated with abnormal cytokine metabolism.
XX
SQ Sequence 21721 BP; 4980 A; 5688 C; 6170 G; 4865 T; 18 other;

Query Match 0.2%; Score 59; DB 20; Length 21721;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9009 ccagtgtaattttgtatttttagtagagatgggtttccaccatgttggccaggctggt 9067
|||||
Db 4077 ccagtgtaattttgtatttttagtagagatgggtttccaccatgttggccaggctggt 4135
|||||

RESULT 30
AAW83426
ID AAX83426 standard; DNA; 22976 BP.
XX
AC AAX83426;
XX
DT 31-AUG-1999 (first entry)
XX
DE Genomic region containing human lipolysis stimulated receptor gene.
XX
KW Human; lipolysis stimulated receptor; LSR; lipoprotein; free fatty acid;
KW cytokine; probe; primer; amplification; hybridisation; detection; leptin;
KW allele; variant; mutation; deletion; loss of heterozygosity; chylomicron;
KW transgenic animal; gene expression; triglyceride; eating disorder;
KW obesity; atheromatosis; atherosclerosis; hypertension; diabetes;
KW anorexia; metabolism; ds.
XX
OS Homo sapiens.
XX
PN WO9907737-A2.
XX
PD 18-FEB-1999.
XX
PF 06-AUG-1998; 98WO-IB01257.
XX
PR 22-APR-1998; 98FR-0005032.
PR 06-AUG-1997; 97FR-0010088.
XX
PA (GEST) GENSET.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bihaïn B, Bougueleret L, Yen-Potin F;
XX
DR WPI; 1999-190035/16.
XX
PT Lipolysis stimulated receptor involved in leptin metabolism - and
PT controlling its activity for treatment of obesity, anorexia or
PT abnormal cytokine metabolism

XX
PS Claim 40; Page 222-242; 279pp; French.
XX
CC This sequence represents the genomic region containing the coding region
CC for the human lipolysis stimulated receptor (LSR). The novel LSR binds
CC lipoproteins in presence of free fatty acids (FFA) and binds cytokines
CC in absence of FFA. The nucleic acid and its fragments are used as
CC probes and primers for detection and/or amplification of LSR genes;
CC for production of recombinant LSR; for detecting allelic variants,
CC mutations, deletions, loss of heterozygosity and genetic abnormalities
CC in the gene. LSR, recombinant cells and transgenic animals are used to
CC screen for chemical interacting with LSR, also to study expression and
CC activity of LSR and its interactions. The chemicals, and leptin, are
CC used to modulate the number of LSR in a cell, its recycling rate and/or
CC specificity of receptor activity, particularly for reducing the level of

CC leptin, lipoproteins, chylomicrons and/or triglycerides. The chemicals
CC are thus useful for treating eating disorders, particularly obesity (and
CC related diseases such as atherosclerosis, atherosclerosis, hypertension
CC and diabetes) or anorexia, also disease associated with abnormal cytokine
CC metabolism.

XX
SQ Sequence 22976 BP; 5120 A; 6191 C; 6677 G; 4988 T; 0 other;

Query Match 0.2%; Score 59; DB 20; Length 22976;
Best Local Similarity 100.08; Pred. No. 1.7e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9009 ccagtgtaatttttttagtagagatgggtttccaccatgttgccaggctggt 9067
|||||

Db 4076 ccagtgtaatttttttagtagagatgggtttccaccatgttgccaggctggt 4134

RESULT 31

AAA50273

ID AAA50273 standard; DNA; 23187 BP.

XX

AC AAA50273;

XX

DT 07-NOV-2000 (first entry)

XX

DE Human lipolysis stimulated receptor (LSR) gene.

XX

KW Lipolysis stimulated receptor; LSR; chromosome 19q13.1; human;

KW single nucleotide polymorphism; biallelic marker; obesity;

KW atherosclerosis; insulin resistance; hypertension; hyperlipidemia;

KW hypertriglyceridemia; cardiovascular disease; microangiopathy;

KW syndrome X; diagnosis; therapy; genotyping; ds.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT 5'UTR 1..2000

FT

FT /*tag= a

FT /note= "potential 5' regulatory region"

FT

FT /*tag= b

FT /number= 1

FT 2357..3539

FT

FT /*tag= c

FT /number= 1

FT 3540..3884

FT

FT /*tag= d

FT /number= 2

FT 3885..12162

FT

FT /*tag= e

FT /number= 2

FT 12163..12282

FT

FT /*tag= f

FT /number= 3

FT 12283..15143

FT

FT /*tag= g

FT /number= 3

FT 15144..15200

FT

FT /*tag= h

FT /number= 4

FT 15201..15764

FT

FT /*tag= i

FT /number= 4

FT 15765..15911

FT

FT /*tag= j

FT /number= 5

FT 15912..19578

FT

FT /*tag= k

FT /number= 5

FT 19579..19752

FT

FT /*tag= l

FT /number= 6

FT

FT intron

FT

FT /*tag= m

FT /number= 6

FT 19899..19958

FT

FT /*tag= n

FT /number= 7

FT 19959..20055

FT

FT /*tag= o

FT /number= 7

FT 20056..20187

FT

FT /*tag= p

FT /number= 8

FT 20188..20328

FT

FT /*tag= q

FT /number= 8

FT 20329..20957

FT

FT /*tag= r

FT /number= 9

FT 20958..21046

FT

FT /*tag= s

FT /number= 9

FT 21047..21187

FT

FT /*tag= t

FT /number= 10

FT replace(818, G)

FT

FT /*tag= u

FT /standard_name= "single nucleotide polymorphism"

FT /note= "marker 17-2-297 (A31)"

FT replace(1243, T)

FT

FT /*tag= v

FT /frequency= 0.15

FT /standard_name= "single nucleotide polymorphism"

FT /note= "marker 9-19-148 (A10)"

FT replace(1374, A)

FT

FT /*tag= w

FT /standard_name= "single nucleotide polymorphism"

FT /note= "marker 9-19-256 (A32)"

FT replace(1401, T)

FT

FT /*tag= x

FT /frequency= 0.12

FT /standard_name= "single nucleotide polymorphism"

FT /note= "marker 9-19-307 (A11)"

FT delete(1535)

FT

FT /*tag= y

FT /standard_name= "single nucleotide polymorphism"

FT /note= "marker 9-19-442 (A12)"

FT replace(1788, A)

FT

FT /*tag= z

FT /standard_name= "single nucleotide polymorphism"

FT /note= "marker 9-20-187 (A13)"

FT replace(2391, G)

FT

FT /*tag= aa

FT /frequency= 0.24

FT /standard_name= "single nucleotide polymorphism"

FT /note= "marker 9-1-308 (A14)"

FT replace(3778, T)

FT

FT /*tag= ab

FT /frequency= 0.29

FT /standard_name= "single nucleotide polymorphism"

FT /note= "marker 9-3-324 (A15)"

FT replace(4498, T)

FT

FT /*tag= ac

FT /frequency= 0.22

FT /standard_name= "single nucleotide polymorphism"

FT /note= "marker 99-14419-424 (A16)"

FT replace(15007, A)

FT

FT /*tag= ad

FT /frequency= 0.35

FT /standard_name= "single nucleotide polymorphism"

FT /note= "marker 9-24-260 (A17)"

FT replace(15233, A)

FT

FT /*tag= ae

FT /frequency= 0.15

FT

FT

```
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-24-486 (A18)"
FT replace(15826,T)
FT /*tag= af
FT /frequency= 0.01
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-6-187 (A19)"
FT replace(19567,G)
FT /*tag= ag
FT /frequency= 0.35
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-7-148 (A20)"
FT replace(19744,A)
FT /*tag= ah
FT /frequency= 0.14
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-7-325 (A21)"
FT replace(19786,C)
FT /*tag= ai
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-7-367 (A22)"
FT replace(20158,G)
FT /*tag= aj
FT /frequency= 0.05
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-9-246 (A23)"
FT delete(20595..20597)
FT /*tag= ak
FT /frequency= 0.26
FT /note= "marker LSRX9-BM (17-1-240) (A24)"
FT replace(21108,G)
FT /*tag= al
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker LSRX10-BM (A25)"
FT replace(606,T)
FT /*tag= am
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'1"
FT insert(5141,G)
FT /*tag= an
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'2"
FT insert(7428,C)
FT /*tag= ao
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'3"
FT replace(8394,G)
FT /*tag= ap
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'4"
FT replace(8704,C)
FT /*tag= aq
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'5"
FT replace(9028,A)
FT /*tag= ar
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'6"
FT delete(9950..9957)
FT /*tag= as
FT replace(9977,C)
FT /*tag= at
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'8"
FT replace(10021,A)
FT /*tag= au
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'9"
FT replace(11878,T)
FT /*tag= av
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'10"
```

```
FT variation delete(19040)
Query Match 0.2% Score 59; DB 21; Length 23187;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9009 cccagtttaatttttttagtagatgggggtttccaccatgttgccaggctggt 9067
|||||
Db 4179 cccagtttaatttttttagtagatgggggtttccaccatgttgccaggctggt 4237

RESULT 32
AAFG2331
ID AAF62331 standard; DNA; 23187 BP.
XX
AC AAF62331;
XX
DT 06-JUN-2001 (first entry)
XX
DE Human leptin fragment coding sequence SEQ ID NO: 1.
XX
KW Leptin; human; LSR; lipolysis stimulated receptor; obesity;
KW hypertension; anorexia; cachexia; stroke; atherosclerosis; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_feature 1..2000
FT /*tag= a
FT /note= "5' regulatory region"
FT primer_bind 523..544
FT /*tag= b
FT misc_binding 587..605
FT /*tag= c
FT allele replace(606,T)
FT /*tag= d
FT misc_binding complement(607..625)
FT /*tag= e
FT misc_binding 799..817
FT /*tag= f
FT allele replace(818,C)
FT /*tag= g
FT misc_binding complement(819..837)
FT /*tag= h
FT primer_bind 946..963
FT /*tag= i
FT primer_bind complement(1047..1068)
FT /*tag= j
FT primer_bind 1096..1115
FT /*tag= k
FT misc_binding 1224..1242
FT /*tag= l
FT allele replace(1243,T)
FT /*tag= m
FT misc_binding complement(1244..1262)
FT /*tag= n
FT misc_binding 1330..1373
FT /*tag= o
FT allele replace(1374,G)
FT /*tag= p
FT misc_binding complement(1375..1393)
FT /*tag= q
FT misc_binding 1382..1400
FT /*tag= r
FT primer_bind complement(1385..1402)
FT /*tag= s
FT allele replace(1401,T)
FT /*tag= t
FT misc_binding complement(1402..1420)
FT /*tag= u
FT misc_binding 1516..1534
FT /*tag= v
FT primer_bind 1602..1621
```

```
FT primer_bind /*tag= w complement(1616..1635)
FT /*tag= x 1769..1787
FT misc_binding /*tag= y replace(1788,C)
FT allele /*tag= z complement(1789..1807)
FT misc_binding /*tag= aa 2001..21190
FT CDS /*tag= ab /*product= "LSR"
FT exon /*note= "this sequence contains introns"
FT 2001..2356
FT /*tag= ac
FT /*number= 1
FT primer_bind 2036..2053
FT /*tag= ad
FT primer_bind 2062..2081
FT /*tag= ae
FT primer_bind complement(2074..2093)
FT /*tag= af
FT primer_bind 2084..2102
FT /*tag= ag
FT misc_binding 2372..2390
FT /*tag= ah
FT allele replace(2391,C)
FT /*tag= ai
FT misc_binding complement(2392..2410)
FT /*tag= aj
FT primer_bind complement(2563..2580)
FT /*tag= ak
FT primer_bind complement(2470..2489)
FT /*tag= al
FT primer_bind complement(2483..2500)
FT /*tag= am
FT primer_bind 3455..3474
FT /*tag= an
FT exon 3540..3884
FT /*tag= ao
FT /*number= 2
FT misc_binding 3759..3777
FT /*tag= ap
FT primer_bind 3775..3792
FT /*tag= aq
FT allele replace(3778,T)
FT /*tag= ar
FT misc_binding complement(3779..3797)
FT /*tag= as
FT primer_bind complement(3882..3901)
FT /*tag= at
FT primer_bind complement(4336..4356)
FT /*tag= au
FT primer_bind 4444..4463
FT /*tag= av
FT allele replace(4498,G)
FT /*tag= aw
FT primer_bind complement(4902..4920)
FT /*tag= ax
FT misc_binding 4979..4997
FT /*tag= ay
FT misc_binding complement(4999..5017)
FT /*tag= az
FT misc_binding 5122..5140
FT /*tag= ba
FT misc_binding complement(5142..5160)
FT /*tag= bb
FT primer_bind 6638..6655
FT /*tag= bc
FT primer_bind complement(7072..7089)
FT /*tag= bd
FT misc_binding 7409..7427
FT /*tag= be
```

```
FT misc_binding complement(7429..7447)
FT /*tag= bf
FT primer_bind 7995..8012
FT /*tag= bg
FT misc_binding 8375..7393
FT /*tag= bh
FT allele replace(8394,C)
FT /*tag= bi
FT misc_binding complement(8395..9413)
FT /*tag= bj
FT primer_bind complement(8576..8593)
FT /*tag= bk
FT misc_binding 8685..8703
FT /*tag= bl
FT allele replace(8704,T)
FT /*tag= bm
FT misc_binding complement(8705..8723)
FT /*tag= bn
FT misc_binding 9009..9027
FT /*tag= bo
FT allele replace(9028,G)
FT /*tag= bp
FT misc_binding complement(9029..9047)
FT /*tag= bq
FT primer_bind 9622..9639
FT /*tag= br
FT misc_binding 9931..9949
FT /*tag= bs
FT misc_binding complement(9951..9969)
FT /*tag= bt
FT misc_binding 9958..9976
FT /*tag= bu
FT primer_bind 9964..9981
FT /*tag= bv
FT allele replace(9977,T)
FT /*tag= bw
FT misc_binding complement(9978..9996)
FT /*tag= bx
FT misc_binding 10002..10020
FT /*tag= by
FT allele replace(10021,G)
FT /*tag= bz
FT misc_binding complement(10022..10040)
FT /*tag= ca
FT primer_bind complement(10023..10040)
FT /*tag= cb
FT primer_bind 10492..10512
FT /*tag= cc
FT primer_bind complement(10546..10563)
FT /*tag= cd
FT misc_binding 11857..11877
FT /*tag= ce
FT primer_bind complement(10996..11015)
FT /*tag= cf
FT allele replace(11878,T)
FT /*tag= cg
FT misc_binding complement(11879..11897)
FT /*tag= ch
FT primer_bind 11972..11990
FT /*tag= ci
FT primer_bind 12005..12023
FT /*tag= cj
FT exon 12163..12282
FT /*tag= ck
FT /*number= 3
FT primer_bind complement(12417..12436)
FT /*tag= cl
```

Query Match 0.2%; Score 59; DB 22; Length 23187;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9009 cccagttaattttgtatttttagtagatggggtttccacctgttgccagctagt 9067

```
Db 4179 cccagtgtaattttagatgggggtttccaccatgttgccacggtggt 4237
|||||
RESULT 33
AAT94101
ID AAT94101 standard; DNA: 53526 BP.
XX AC AAT94101;
XX DT 01-JUN-1998 (first entry)
XX DE Human PKD1 gene.
XX KW Human; polycystic kidney disease 1; PKD1; treatment;
XX KW autosomal dominant polycystic kidney disease; APKD; ss.
XX OS Homo sapiens.
XX PN WO9744457-A1.
XX PD 27-NOV-1997.
XX PF 22-MAY-1997; 97WO-US08799.
XX PR 03-JUN-1996; 96US-0658136.
XX PR 24-MAY-1996; 96US-0655360.
XX PA (GENZ ) GENZYME CORP.
XX PI Burn T, Connors T, Dackowski W, Germino G, Klingler K;
XX PI Qian F;
XX DR WPI; 1998-018511/02.
XX PT Human polycystic kidney disease gene, PKD1 - useful to treat and
XX PT diagnose human autosomal or adult onset polycystic kidney disease
XX PS Claim 2; Pages 90-118; 257pp; English.
XX CC The present sequence is the human polycystic kidney disease 1
XX CC (PKD1) gene. The PKD1 gene or polypeptide may be used to treat
XX CC autosomal dominant polycystic kidney disease (APKD), and identify
XX CC carriers of mutant PKD1 genes, i.e. subjects susceptible to APKD.
XX CC Antibodies (Ab) that distinguish between normal and mutant PKD1
XX CC sequences can also be used in diagnostic tests. Anti-PKD1 Ab may
XX CC also be used to perform subcellular and histochemical localisation
XX CC studies, and to block the function of PKD1. Ab are also useful in
XX CC rational drug design studies to identify and test inhibitors of
XX CC PKD1. Sense and antisense sequences derived from the PKD1 gene may
XX CC used for detection and therapy.
XX SQ Sequence 53526 BP; 8486 A; 17665 C; 15768 G; 11607 T; 0 other;

Query Match 0.2%; Score 59; DB 19; Length 53526;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12652 agagacggggtttctccacgttggtcaggtggtctcaaaactctgacctcagtgatc 12710
|||||
Db 7254 agagacggggtttctccacgttggtcaggtggtctcaaaactctgacctcagtgatc 7312

RESULT 34
AAT18551
ID AAT18551 standard; DNA: 53577 BP.
XX AC AAT18551;
XX DT 06-MAY-1997 (first entry)
XX DE Human polycystic kidney disease normal PKD1 gene.
```

```
XX Adult onset polycystic kidney disease; APKD; autosomal dominant;
KW mutant; transversion; transition; deletion; insertion; ds.
XX OS Homo sapiens.
XX Key Location/Qualifiers
XX misc_feature 4379..5272
XX FT /*tag= a
XX FT /note= "specifically claimed region of intronless
FT cDNA identified by exon trapping"
XX FT old_sequence replace(50652..50653, cg)
XX FT /*tag= b
XX FT /note= "changes Val codon to Leu codon"
XX FT old_sequence replace(50796..50797, cg)
XX FT /*tag= c
XX FT /note= "replaces Val codon by Leu codon"
XX FT old_sequence insert(51827..51828, cc)
XX FT /*tag= d
XX FT /note= "insertion, results in frameshift"
XX PN WO9612033-A1.
XX XX
XX PD 25-APR-1996.
XX PF 11-OCT-1995; 95WO-US13357.
XX PR 31-JAN-1995; 95US-0381520.
XX PR 12-OCT-1994; 94US-0323443.
XX XX
XX PA (ICIG-) IG LAB INC.
XX PI (UWJO ) UNIV JOHNS HOPKINS.
XX PI Burn TC, Connors TD, Dackowski W, Germino G, Klingler KW;
XX PI Landes GM, Qian F;
XX DR WPI; 1996-222017/22.
XX XX
XX PT Isolated human polycystic kidney disease gene and its mutants -
XX PT useful for treatment of polycystic kidney disease and screening for
XX PT carriers
XX PS Claim 1; Fig 1; 65pp; English.
XX CC The present sequence is that of the normal human PKD1 gene from
XX CC chromosome 16. Mutations in this gene (e.g. transitions,
XX CC transversions, deletions and/or insertions) are associated with
XX CC adult-onset polycystic kidney disease (APKD). The PKD1 locus is
XX CC GC-rich (62.4%). Comparison of this sequence with a previously
XX CC reported partial cDNA sequence revealed differences at three
XX CC locations (see features table). The most significant difference is
XX CC the presence of two additional cytosine residues on the plus-strand
XX CC at position 4566 of the previously reported sequence. The insertion
XX CC results in a frame-shift in the predicted protein coding sequence,
XX CC leading to replacement of 92 C-terminal amino acids with a novel
XX CC 12 amino acid C-terminus. The PKD1 gene contains 23 Alu repeats.
XX CC There is a region consisting of 17 tandem copies of a perfect 27 bp
XX CC repeat and two large CT-rich regions.
XX SQ Sequence 53577 BP; 8495 A; 17681 C; 15785 G; 11616 T; 0 other;

Query Match 0.2%; Score 59; DB 17; Length 53577;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12652 agagacggggtttctccacgttggtcaggtggtctcaaaactctgacctcagtgatc 12710
|||||
Db 7254 agagacggggtttctccacgttggtcaggtggtctcaaaactctgacctcagtgatc 7312

RESULT 35
AAT94108
```


ID AAT94108 standard; DNA; 53577 BP.
XX
AC AAT94108;
XX
XX
DT 01-JUN-1998 (first entry)
XX
DE Human PKD1 locus between chromosomal markers ATPL (ATP6C) and D16S84.
XX
XX Human; polycystic kidney disease 1; PKD1; treatment;
KW autosomal dominant polycystic kidney disease; APKD; ss.
XX
XX Homo sapiens.
OS
XX WO9744457-A1.
PN
XX
PD 27-NOV-1997.
XX
XX
PF 22-MAY-1997; 97WO-US08799.
XX
XX 03-JUN-1996; 96US-0658136.
PR 24-MAY-1996; 96US-0655360.
XX
XX (GENZ) GENZYME CORP.
PA
XX Burn T, Connors T, Dackowski W, Germino G, Klingner K;
PI Qian F;
XX
XX WPI; 1998-018511/02.
DR
XX Human polycystic kidney disease gene, PKD1 - useful to treat and
PT diagnose human autosomal or adult onset polycystic kidney disease
PT
XX
PS Example 5; Pages 60-89; 257pp; English.
XX
XX The present sequence is the human polycystic kidney disease 1
CC (PKD1) locus between chromosomal markers ATPL (ATP6C) and D16S84.
CC The PKD1 gene or polypeptide may be used to treat autosomal
CC dominant polycystic kidney disease (APKD), and identify carriers
CC of mutant PKD1 genes, i.e. subjects susceptible to APKD. Antibodies
CC (Ab) that distinguish between normal and mutant PKD1 sequences can
CC also be used in diagnostic tests. Anti-PKD1 Ab may also be used to
CC perform subcellular and histochemical localisation studies, and to
CC block the function of PKD1. Ab are also useful in rational drug
CC design studies to identify and test inhibitors of PKD1. Sense and
CC antisense sequences derived from the PKD1 gene may be used for
CC detection and therapy.
XX
SQ Sequence 53577 BP; 8495 A; 17684 C; 15782 G; 11616 T; 0 other;

Query Match 0.2%; Score 59; DB 19; Length 53577;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12652 agagacggggtttccacgttggtcaggtggtctcaaacctcgtacctcaggtgac 12710
|||||
Db 7253 agagacggggtttccacgttggtcaggtggtctcaaacctcgtacctcaggtgac 7311

RESULT 36
AAZ98085/c
ID AAZ98085 standard; cDNA; 2191 BP.
XX
AC AAZ98085;
XX
DT 09-MAY-2000 (first entry)
XX
XX Human secreted protein encoding nucleotide sequence SEQ ID NO:79.
DE
XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antiinflammatory; neurotropic; neuroprotective; antiallergic; cancer;
KW tumour; neurodegenerative disorder; developmental abnormality; allergy;
KW foetal deficiency; blood disorder; immune system disorder; arthritis;

KW autoimmune disease; hepatic disease; renal disease; inflammation;
KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;
KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;
KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;
KW reproductive disorder; gastrointestinal disorder; respiratory disorder;
KW metabolic disorder; food additive; preservative; ss.
XX
OS Homo sapiens.
XX
PN WO200004140-A1.
XX
PD 27-JAN-2000.
XX
XX 14-JUL-1999; 99WO-US15849.
PF
XX 15-JUL-1998; 98US-0092921.
PR 15-JUL-1998; 98US-0092922.
PR 15-JUL-1998; 98US-0092956.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;
PI Mucenski M, Endress GA, Soppet DR;
XX
DR WPI; 2000-161128/14.
DR P-PSDB; AAY87132.
XX
XX New isolated human genes, useful for diagnosis and treatment of, e.g.
PT cancers, neurological or blood disorders -
PT
XX
PS Claim 1; Page 353-354; 494pp; English.
XX
XX The polynucleotide sequences given in AAY98017 to AAZ98108 encode the
CC human secreted proteins given in AAY87064 to AAY87223. Human secreted
CC protein can have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: cytostatic;
CC immunosuppressive; antiinflammatory; neurotropic; neuroprotective; and
CC antiallergic. The polynucleotides and their corresponding secreted
CC polypeptides are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new
CC polynucleotides. Human secreted protein s and their polynucleotides can
CC be used for developing products for the diagnosis or treatment of cancer,
CC tumours, neurodegenerative disorders, developmental abnormalities and
CC foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, inflammation,
CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC cardiovascular disorders, reproductive disorders, gastrointestinal
CC disorders, respiratory disorders and metabolic disorders. The
CC proteins or polynucleotides can also be used as food additives or
CC preservatives. The proteins are also useful for identifying their
CC binding partners. AAY98008 to AAZ98016 and AAY87063 are sequence used in
CC the exemplification of the present invention.
XX
SQ Sequence 2191 BP; 597 A; 524 C; 503 G; 565 T; 2 other;

Query Match 0.2%; Score 58; DB 21; Length 2191;
Best Local Similarity 100.0%; Pred. No. 5.7e-10;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17555 caggctgggtctgaactcctgacctcaggtgacccaccacccacccagctcccaagtg 17612
|||||
Db 1983 CAGGCTGGTCTCGAACTCTGACCTCAGGTGATCCACCCACCTCAGCCTCCCAAGTG 1926

RESULT 37
AAT72165/c
ID AAT72165 standard; cDNA to mRNA; 2688 BP.


```
RESULT 39
AAD02697/c
ID AAD02697 standard; DNA; 160552 BP.
XX AC AAD02697;
XX DT
XX DE
XX DE Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.
XX KW Human; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1; ds.
XX OS Homo sapiens.
XX FH
XX FH Key Location/Qualifiers
FT exon 32847..32922
FT FT /*tag= a
FT FT /number= 1
FT FT /label= 4a_504
FT FT 32923..35592
FT FT /*tag= b
FT FT /cons_splice= (5'site:NO, 3'site:YES)
FT FT 35593..35674
FT FT /*tag= c
FT FT /number= 2
FT FT /label= 4a_503
FT FT 35675..45093
FT FT /*tag= d
FT FT 45094..45185
FT FT /*tag= e
FT FT /number= 3
FT FT /label= 4a_502
FT FT 45186..46633
FT FT /*tag= f
FT FT /cons_splice= (5'site:NO, 3'site:NO)
FT FT 46634..46700
FT FT /*tag= g
FT FT /number= 4
FT FT /label= 4a_501
FT FT 46701..47938
FT FT /*tag= h
FT FT /cons_splice= (5'site:YES, 3'site:NO)
FT FT 47939..49746
FT FT /*tag= i
FT FT /number= 5
FT FT /note= "Includes 17 base pairs of 5'UTR, the ORF
FT and all of 3'UTR"
FT FT 47939..47955
FT FT /*tag= j
FT FT /note= "Portion of 5' untranslated region (5'UTR)"
FT FT 47956..49128
FT FT /*tag= k
FT FT /product= "Human glycosyl transferase-4alpha
FT (GST-4alpha)"
FT FT 49129..49746
FT FT /*tag= l
FT FT 83257..83347
FT FT /*tag= m
FT FT /label= 4a_502
FT FT 83348..96412
FT FT /*tag= n
FT FT /cons_splice= (5'site:NO, 3'site:NO)
FT FT 96413..96484
```

```
FT FT /*tag= o
FT FT /label= 4a_501
FT FT 96485..98456
FT FT /*tag= p
FT FT /cons_splice= (5'site:NO, 3'site:NO)
FT FT 98457..99968
FT FT /*tag= q
FT FT /note= "Includes 17 base pairs of 5'UTR, the ORF
FT and all of 3'UTR"
FT FT 98457..98473
FT FT /*tag= r
FT FT /note= "Portion of 5' untranslated region (5'UTR)"
FT FT 98474..99661
FT FT /*tag= s
FT FT /product= "Human glycosyl transferase-4beta
FT (GST-4beta)"
FT FT 99662..99968
FT FT /*tag= t
XX WO200106015-A1.
XX 25-JAN-2001.
XX 19-JUL-2000; 2000WO-US19741.
XX 20-JUL-1999; 990US-0144694.
XX 13-JUL-2000; 2000US-0593828.
XX (REGC ) UNIV CALIFORNIA.
XX Rosen SD, Lee JK, Hemmerich S;
XX WPI; 2001-138471/14.
XX P-PSDB; AAY72639, AAY72640.
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications -
XX Example 1; Page 62-104; 128pp; English.
XX The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic
XX DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on
XX chromosome 16q23.1.
XX GST is a type 2 membrane protein useful for inhibiting a binding event
XX between a selectin and a selectin ligand, which comprises contacting the
XX selectin with a non-sulphated selectin ligand, GST and a small molecular
XX agent that inhibits the sulphation activity of GST. GST is also useful
XX in inhibiting a selectin mediated binding event. GST is useful in gene
XX therapy to treat disorders such as acute or chronic inflammation,
XX systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
XX nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX during transplantation.
XX SQ Sequence 160552 BP; 40281 A; 37573 C; 38015 G; 44564 T; 119 other;
```

```
Query Match 0.2%; Score 58; DB 22; Length 160552;
Best Local Similarity 100.0%; Pred. NO. 2.4e-10;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 11123 ggcataattttgttttagtagacggggtttccaccattgtggccaggctggtct 11180
|||||
Db 27396 GGCCTAAATTTTGTCTTTTACTAGACAGGGGTTTCACCATGTTGCCAGCGTGGTCT 27339
```

```
RESULT 40
AAC98168
```



```
AC AAC72064;
XX
XX 09-FEB-2001 (first entry)
XX
XX Single nucleotide polymorphism containing sequence #631.
XX
XX Single nucleotide polymorphism; SNP; human; genetic disease;
KW disease susceptibility; cardiovascular system; endocrine system;
KW neurological system; forensic testing; paternity testing; ds.
XX
XX Homo sapiens.
XX
XX WO200058519-A2.
XX
XX 05-OCT-2000.
XX
XX 30-MAR-2000; 2000WO-US08440.
XX
XX 31-MAR-1999; 99US-0127248.
XX
XX (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX (AFFY-) AFFYMETRIX INC.
XX
XX Altshuler D, Cargill M, Daley GQ, Ireland JS, Lander ES;
PI Lipshutz RJ, Patil N, Sklar P;
XX
XX WPI; 2000-611722/58.
XX
XX Nucleic acid selected from one of 106 genes comprising single
XX nucleotide polymorphisms, allele-specific oligonucleotides to the genes
XX are useful for phenotypic correlations, forensics, paternity testing,
XX medicine and genetic analysis -
XX
XX Claim 1; Fig 5; 214pp; English.
XX
XX The present invention is concerned with a number of human single
XX nucleotide polymorphisms (SNPs) which the inventors identified in human
XX genes. These SNPs can be used in disease diagnosis and prediction of an
XX individual's susceptibility to disease, in forensic and paternity testing
XX and in genetic mapping. In particular, the SNPs of the invention can be
XX used to diagnose susceptibility to diseases of the cardiovascular,
XX endocrine and neurological systems, such as coronary artery disease,
XX schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
XX diseases.
XX
XX Note: The degenerate codon within the sequence represents the position
XX of an SNP, for example the letter S represents a polymorphism where the
XX nucleotide may be C or G.
XX
XX Sequence 792 BP; 213 A; 195 C; 145 G; 238 T; 1 other;

Query Match 0.2%; Score 57; DB 21; Length 792;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4588 ccaggctgagtcgagtcgcatgctcgcctcactgcacccacccctcctgggtt 4644
|||||
Db 213 ccaggctgagtcgagtcgcatgctcgcctcactgcacccacccctcctgggtt 269

RESULT 43
AAC72079
ID AAC72079 standard; DNA; 792 BP.
XX
XX AAC72079;
XX
XX 09-FEB-2001 (first entry)
XX
XX Single nucleotide polymorphism containing sequence #636.
XX
XX Single nucleotide polymorphism; SNP; human; genetic disease;
KW disease susceptibility; cardiovascular system; endocrine system;
KW neurological system; forensic testing; paternity testing; ds.
XX
```

```
XX Homo sapiens.
XX
XX WO200058519-A2.
XX
XX 05-OCT-2000.
XX
XX 30-MAR-2000; 2000WO-US08440.
XX
XX 31-MAR-1999; 99US-0127248.
XX
XX (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX (AFFY-) AFFYMETRIX INC.
XX
XX Altshuler D, Cargill M, Daley GQ, Ireland JS, Lander ES;
PI Lipshutz RJ, Patil N, Sklar P;
XX
XX WPI; 2000-611722/58.
XX
XX Nucleic acid selected from one of 106 genes comprising single
XX nucleotide polymorphisms, allele-specific oligonucleotides to the genes
XX are useful for phenotypic correlations, forensics, paternity testing,
XX medicine and genetic analysis -
XX
XX Claim 1; Fig 5; 214pp; English.
XX
XX The present invention is concerned with a number of human single
XX nucleotide polymorphisms (SNPs) which the inventors identified in human
XX genes. These SNPs can be used in disease diagnosis and prediction of an
XX individual's susceptibility to disease, in forensic and paternity testing
XX and in genetic mapping. In particular, the SNPs of the invention can be
XX used to diagnose susceptibility to diseases of the cardiovascular,
XX endocrine and neurological systems, such as coronary artery disease,
XX schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
XX diseases.
XX
XX Note: The degenerate codon within the sequence represents the position
XX of an SNP, for example the letter S represents a polymorphism where the
XX nucleotide may be C or G.
XX
XX Sequence 792 BP; 213 A; 195 C; 144 G; 239 T; 1 other;

Query Match 0.2%; Score 57; DB 21; Length 792;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4588 ccaggctgagtcgagtcgcatgctcgcctcactgcacccacccctcctgggtt 4644
|||||
Db 213 ccaggctgagtcgagtcgcatgctcgcctcactgcacccacccctcctgggtt 269

RESULT 44
AAC72094
ID AAC72094 standard; DNA; 792 BP.
XX
XX AAC72094;
XX
XX 09-FEB-2001 (first entry)
XX
XX Single nucleotide polymorphism containing sequence #641.
XX
XX Single nucleotide polymorphism; SNP; human; genetic disease;
KW disease susceptibility; cardiovascular system; endocrine system;
KW neurological system; forensic testing; paternity testing; ds.
XX
XX Homo sapiens.
XX
XX WO200058519-A2.
XX
XX 05-OCT-2000.
XX
XX 30-MAR-2000; 2000WO-US08440.
XX
```

```
PR 31-MAR-1999; 99US-0127248.
XX (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA (AFFY-) AFFYMETRIX INC.
XX
XX Altshuler D, Cargill M, Daley GQ, Ireland JS, Lander ES;
PI Lipshutz RJ, Patil N, Sklar P;
XX
XX WPI; 2000-611722/58.
XX
XX Nucleic acid selected from one of 106 genes comprising single
XX nucleotide polymorphisms, allele-specific oligonucleotides to the genes
XX are useful for phenotypic correlations, forensics, paternity testing,
XX medicine and genetic analysis -
XX
XX Claim 1; Fig 5; 214pp; English.
XX
XX The present invention is concerned with a number of human single
XX nucleotide polymorphisms (SNPs) which the inventors identified in human
XX genes. These SNPs can be used in disease diagnosis and prediction of an
XX individual's susceptibility to disease, in forensic and paternity testing
XX and in genetic mapping. In particular, the SNPs of the invention can be
XX used to diagnose susceptibility to diseases of the cardiovascular,
XX endocrine and neurological systems, such as coronary artery disease,
XX schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
XX diseases.
XX Note: The degenerate codon within the sequence represents the position
XX of an SNP, for example the letter S represents a polymorphism where the
XX nucleotide may be C or G.
XX
XX Sequence 792 BP; 213 A; 195 C; 144 G; 239 T; 1 other;
SQ

Query Match 0.2%; Score 57; DB 21; Length 792;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4588 ccaggctggagtcagtgccatgatctcggctcactcgaacctccacctctgggtt 4644
Db 213 ccaggctggagtcagtgccatgatctcggctcactcgaacctccacctctgggtt 269

RESULT 45
AAC72106
ID AAC72106 standard; DNA; 792 BP.
XX
XX AAC72106;
AC
AC AAC72106;
DT 09-FEB-2001 (first entry)
XX
XX Single nucleotide polymorphism containing sequence #645.
XX
XX Single nucleotide polymorphism; SNP; human; genetic disease;
XX disease susceptibility; cardiovascular system; endocrine system;
XX neurological system; forensic testing; paternity testing; ds.
XX
XX Homo sapiens.
XX
XX WO200058519-A2.
XX
XX 05-OCT-2000.
XX
XX 30-MAR-2000; 2000WO-US08440.
XX
XX 31-MAR-1999; 99US-0127248.
XX
XX (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX (AFFY-) AFFYMETRIX INC.
XX
XX Altshuler D, Cargill M, Daley GQ, Ireland JS, Lander ES;
PI Lipshutz RJ, Patil N, Sklar P;
XX
XX WPI; 2000-611722/58.
XX
XX Nucleic acid selected from one of 106 genes comprising single
XX nucleotide polymorphisms, allele-specific oligonucleotides to the genes
XX are useful for phenotypic correlations, forensics, paternity testing,
XX medicine and genetic analysis -
XX
XX Claim 1; Fig 5; 214pp; English.
XX
XX The present invention is concerned with a number of human single
XX nucleotide polymorphisms (SNPs) which the inventors identified in human
XX genes. These SNPs can be used in disease diagnosis and prediction of an
XX individual's susceptibility to disease, in forensic and paternity testing
XX and in genetic mapping. In particular, the SNPs of the invention can be
XX used to diagnose susceptibility to diseases of the cardiovascular,
XX endocrine and neurological systems, such as coronary artery disease,
XX schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
XX diseases.
XX Note: The degenerate codon within the sequence represents the position
XX of an SNP, for example the letter S represents a polymorphism where the
XX nucleotide may be C or G.
XX
XX Sequence 792 BP; 213 A; 195 C; 144 G; 239 T; 1 other;
SQ

Query Match 0.2%; Score 57; DB 21; Length 792;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4588 ccaggctggagtcagtgccatgatctcggctcactcgaacctccacctctgggtt 4644
Db 213 ccaggctggagtcagtgccatgatctcggctcactcgaacctccacctctgggtt 269

RESULT 46
AAC72109
ID AAC72109 standard; DNA; 792 BP.
XX
XX AAC72109;
AC
AC AAC72109;
DT 09-FEB-2001 (first entry)
XX
XX Single nucleotide polymorphism containing sequence #646.
XX
XX Single nucleotide polymorphism; SNP; human; genetic disease;
XX disease susceptibility; cardiovascular system; endocrine system;
XX neurological system; forensic testing; paternity testing; ds.
XX
XX Homo sapiens.
XX
XX WO200058519-A2.
XX
XX 05-OCT-2000.
XX
XX 30-MAR-2000; 2000WO-US08440.
XX
XX 31-MAR-1999; 99US-0127248.
XX
XX (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX (AFFY-) AFFYMETRIX INC.
XX
XX Altshuler D, Cargill M, Daley GQ, Ireland JS, Lander ES;
PI Lipshutz RJ, Patil N, Sklar P;
XX
XX WPI; 2000-611722/58.
XX
XX Nucleic acid selected from one of 106 genes comprising single
XX nucleotide polymorphisms, allele-specific oligonucleotides to the genes
XX are useful for phenotypic correlations, forensics, paternity testing,
XX medicine and genetic analysis -
XX
XX Claim 1; Fig 5; 214pp; English.
XX
XX The present invention is concerned with a number of human single
```

CC nucleotide polymorphisms (SNPs) which the inventors identified in human
CC genes. These SNPs can be used in disease diagnosis and prediction of an
CC individual's susceptibility to disease, in forensic and paternity testing
CC and in genetic mapping. In particular, the SNPs of the invention can be
CC used to diagnose susceptibility to diseases of the cardiovascular,
CC endocrine and neurological systems, such as coronary artery disease,
CC schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
CC diseases.
CC Note: The degenerate codon within the sequence represents the position
CC of an SNP, for example the letter S represents a polymorphism where the
CC nucleotide may be C or G.
XX
SQ Sequence 792 BP; 213 A; 195 C; 145 G; 238 T; 1 other;

Query Match 0.2%; Score 57; DB 21; Length 792;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4588 ccaggctggagtcagtgccatgctcgcctcactgcacccctccactctgggtt 4644
|||||
DB 213 ccaggctggagtcagtgccatgctcgcctcactgcacccctccactctgggtt 269

RESULT 47
AAC72112
ID AAC72112 standard; DNA; 792 BP.
AC AAC72112;
XX
XX
DT 09-FEB-2001 (first entry)
XX
DE Single nucleotide polymorphism containing sequence #647.
XX
KW Single nucleotide polymorphism; SNP; human; genetic disease;
KW disease susceptibility; cardiovascular system; endocrine system;
KW neurological system; forensic testing; paternity testing; ds.
OS Homo sapiens.
XX
XX WO200058519-A2.
XX
PD 05-OCT-2000.
XX
XX 30-MAR-2000; 2000WO-US08440.
XX
XX 31-MAR-1999; 99US-0127248.
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
PA (AFFY-) AFFYMETRIX INC.
PI Altshuler D, Cargill M, Daley GQ, Ireland JS, Lander ES;
PI Lipshutz RJ, Patil N, Sklar P;
XX
XX WPI; 2000-611722/58.
XX
XX Nucleic acid selected from one of 106 genes comprising single
PT nucleotide polymorphisms, allele-specific oligonucleotides to the genes
PT are useful for phenotypic correlations, forensics, paternity testing,
PT medicine and genetic analysis -
XX
PS Claim 1; Fig 5; 214pp; English.
XX
CC The present invention is concerned with a number of human single
CC nucleotide polymorphisms (SNPs) which the inventors identified in human
CC genes. These SNPs can be used in disease diagnosis and prediction of an
CC individual's susceptibility to disease, in forensic and paternity testing
CC and in genetic mapping. In particular, the SNPs of the invention can be
CC used to diagnose susceptibility to diseases of the cardiovascular,
CC endocrine and neurological systems, such as coronary artery disease,
CC schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
CC diseases.
CC Note: The degenerate codon within the sequence represents the position
CC of an SNP, for example the letter S represents a polymorphism where the
CC nucleotide may be C or G.

CC of an SNP, for example the letter S represents a polymorphism where the
CC nucleotide may be C or G.
XX
SQ Sequence 792 BP; 212 A; 195 C; 145 G; 239 T; 1 other;
Query Match 0.2%; Score 57; DB 21; Length 792;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4588 ccaggctggagtcagtgccatgctcgcctcactgcacccctccactctgggtt 4644
|||||
DB 213 ccaggctggagtcagtgccatgctcgcctcactgcacccctccactctgggtt 269

RESULT 48
AAC72115
ID AAC72115 standard; DNA; 792 BP.
XX
AC AAC72115;
XX
XX
DT 09-FEB-2001 (first entry)
XX
DE Single nucleotide polymorphism containing sequence #648.
XX
KW Single nucleotide polymorphism; SNP; human; genetic disease;
KW disease susceptibility; cardiovascular system; endocrine system;
KW neurological system; forensic testing; paternity testing; ds.
OS Homo sapiens.
XX
XX WO200058519-A2.
XX
PD 05-OCT-2000.
XX
XX 30-MAR-2000; 2000WO-US08440.
XX
XX 31-MAR-1999; 99US-0127248.
XX
XX (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX (AFFY-) AFFYMETRIX INC.
PI Altshuler D, Cargill M, Daley GQ, Ireland JS, Lander ES;
PI Lipshutz RJ, Patil N, Sklar P;
XX
XX WPI; 2000-611722/58.
XX
XX Nucleic acid selected from one of 106 genes comprising single
PT nucleotide polymorphisms, allele-specific oligonucleotides to the genes
PT are useful for phenotypic correlations, forensics, paternity testing,
PT medicine and genetic analysis -
XX
PS Claim 1; Fig 5; 214pp; English.
XX
CC The present invention is concerned with a number of human single
CC nucleotide polymorphisms (SNPs) which the inventors identified in human
CC genes. These SNPs can be used in disease diagnosis and prediction of an
CC individual's susceptibility to disease, in forensic and paternity testing
CC and in genetic mapping. In particular, the SNPs of the invention can be
CC used to diagnose susceptibility to diseases of the cardiovascular,
CC endocrine and neurological systems, such as coronary artery disease,
CC schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
CC diseases.
CC Note: The degenerate codon within the sequence represents the position
CC of an SNP, for example the letter S represents a polymorphism where the
CC nucleotide may be C or G.
XX
SQ Sequence 792 BP; 213 A; 195 C; 145 G; 238 T; 1 other;
Query Match 0.2%; Score 57; DB 21; Length 792;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
AAT02714
ID AAT02714 standard; cDNA; 1559 BP.
AC AAT02714;
XX
DT 20-APR-1996 (first entry)
XX
DE MART-1 melanoma antigen.
XX
KW MART-1; melanoma antigen recognised by T-cells; melanoma;
KW metastatic melanoma; tumour-associated antigen; immunogen;
KW diagnosis; prognosis; prophylaxis; therapy; vaccine; ds.
XX
OS Mammalian.
XX
FH Key Location/Qualifiers
FT CDS 54..410
FT /*tag= a
XX
PN W09529193-A2.
XX
PD 02-NOV-1995.
XX
PF 21-APR-1995; 95WO-US05063.
XX
PR 05-APR-1995; 95US-0417174.
PR 22-APR-1994; 94US-0231565.
XX
PA (USSH ) US SEC DEPT HEALTH.
XX
PI Kawakami Y, Rosenberg SA;
XX
DR WPI; 1995-382963/49.
DR P-PSDB; AAR84212.
XX
PT DNA encoding melanoma antigens recognised by T-lymphocytes - also
PT vectors, host cells and antibodies, used to detect, treat and
PT immunise animal against melanoma.
XX
PS Claim 2; Page 115-116; 184pp; English.
XX
CC The nucleic acid encodes a melanoma antigen (MART-1) which is
CC recognized by T-lymphocytes. It is used for recombinant protein
CC production, preferably using a baculo virus vector for expression in
CC insect cell cultures. MART-1 protein is a source of immunogenic
CC peptides (see AAR84196 for peptide M9-2) which are optionally modified
CC (see AAR84783-R84800) and used in medicaments for the treatment or
CC prevention (by immunization) of melanoma. Antibodies against MART-1
CC and its immunogenic peptides may be used in the detection and
CC isolation of MART-1 from a sample, the detection of which is
CC indicative of a disease state (melanoma or metastatic melanoma).
XX
SQ Sequence 1559 BP; 470 A; 330 C; 324 G; 435 T; 0 other;

Query Match 0.2%; Score 57; DB 16; Length 1559;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12671 gttgttcaggctggtctcaaacctcctgacctcagggtgatctgccgcctcagctccc 12727
|||||
Db 1162 gttgttcaggctggtctcaaacctcctgacctcagggtgatctgccgcctcagctccc 1218

RESULT 52
AAZ07987
ID AAZ07987 standard; DNA; 1559 BP.
XX
AC AAZ07987;
XX
DT 10-JAN-2000 (first entry)
XX
DE Human melanoma antigen hMART1 encoding DNA.

Immune response; self-antigen; immune effector cell; cancer; melanoma;
human; melanoma antigen; MART1; ss.
Homo sapiens.
W09946988-A1.
23-SEP-1999.
19-MAR-1999; 99WO-US06034.
20-MAR-1998; 98US-0078890.
(GENZ ) GENZYME CORP.
Nicolette CA;
WPI; 1999-580277/49.
P-PSDB; AAY42634.
Method of inducing an immune reaction to a self-antigen by
administering the antigen, especially useful for treating cancer or
melanoma
Disclosure; Fig 3A-B; 70pp; English.
The invention provides a method of inducing a prophylactic immune
response to a self-antigen in a subject. The method comprises
administering the antigen or its derivative or administering educated
immune effector cells able to recognize and lyse cells expressing the
self-antigen or its derivative. The method is used to stimulate an immune
response against a self-antigen especially one expressed in a cancer or
melanoma. The present sequence represents the DNA sequence encoding
human melanoma antigen hMART1.
Sequence 1559 BP; 470 A; 330 C; 324 G; 435 T; 0 other;

Query Match 0.2%; Score 57; DB 20; Length 1559;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12671 gttgttcaggctggtctcaaacctcctgacctcagggtgatctgccgcctcagctccc 12727
|||||
Db 1162 gttgttcaggctggtctcaaacctcctgacctcagggtgatctgccgcctcagctccc 1218

RESULT 53
AAZ20065
ID AAZ20065 standard; DNA; 1559 BP.
XX
AC AAZ20065;
XX
DT 21-DEC-1999 (first entry)
XX
DE Human MART1 melanocyte differentiation antigen coding region.
XX
KW MART1; melanocyte differentiation antigen; melanoma; human;
KW antigen presentation; adoptive immunotherapy; cancer; therapy;
KW vaccine; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 54..410
FT /*tag= a
XX
PN W09947102-A2.
XX
PD 23-SEP-1999.
XX
PF 19-MAR-1999; 99WO-US06031.
```


PS Claim 9; Page 728; 1046pp; English.

XX The present sequence is one of 3351 sequences in a library of human

CC polynucleotides. The library is used to detect differentially expressed

CC genes correlated with a cancerous state of a mammalian cell and can

CC detect colon, prostate, breast and lung cancer. The library can be used

CC to produce probes for detection of mRNA and to produce additional copies

CC of the polynucleotides. The probes can be used for chromosome mapping of

CC the polynucleotide and for detection of transcription levels. Ribozymes

CC or antisense oligonucleotides can be generated. The polynucleotides and

CC their gene products are used as genetic or biochemical markers (e.g. in

CC blood or tissues) that will detect the earliest changes along the

CC carcinogenesis pathway and/or monitor the efficacy of therapies and

CC preventive interventions. The polynucleotides, polypeptides and

CC antibodies against them can be used in pharmaceutical compositions to

CC treat the cancers and proliferative disorders such as neoplasia,

XX dysplasia and hyperplasia.

SQ Sequence 314 BP; 95 A; 66 C; 81 G; 71 T; 1 other;

Query Match 0.2%; Score 56; DB 22; Length 314;

Best Local Similarity 100.0%; Pred. No. 3.8e-09;

Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9012 agttaatttttttagtagagatgggtttccaccatgttgccagctggt 9067

DB 268 AGTTAATTTTGTATTTTATAGATGGGTTTACCAATGTGGCCAGCTGGT 213

RESULT 56

AAF97863

ID AAF97863 standard; DNA; 11967 BP.

XX AAF97863;

XX 31-MAY-2001 (first entry)

XX Human neuroblastoma cell line NB-1 lp36 nucleotide sequence SEQ ID NO:77.

XX Human; chromosome 1; lp36; neuroblastoma cell line; NB-1; anticancer;

KW tumour suppressor; human lp36 homozygosity deletion domain; tumour;

KW diagnosis; ds.

XX Homo sapiens.

XX WO200116311-A1.

XX 08-MAR-2001.

XX 31-AUG-2000; 2000WO-JP05930.

XX 31-AUG-1999; 99JP-0245962.

PR 09-MAY-2000; 2000JP-0136266.

XX (HISM) HISAMITSU PHARM CO LTD.

PA (CHIB-) CHIBA PREFECTURE.

XX Nakagawara A;

XX WPI; 2001-226686/23.

XX Human lp36 homozygosity deletion domain from the 36-position of first

PT chromosome short arm in human neuroblastoma cell lines, applicable e.g.

PT in gene diagnosis of tumors as well as in developing anti-cancer drugs

XX Example 8; Page 158-163; 226pp; Japanese.

XX The present invention describes a homozygosity deletion domain

CC co-existing in the 36-position of the first chromosome short arm (lp36)

CC in human neuroblastoma. Also described are base sequences from the lp36

CC position of human neuroblastoma cell lines (NB-1 and MASS-NB-SCH-1),

CC Which are tumour suppressor genes in human neuroblastoma. The genes are

CC tumour suppressor genes, base sequence data of which are applicable as

CC tumour markers and reagents in studying mechanism of tumour body

CC formation, and gene diagnosis of tumours as well as in developing

CC anti-cancer drugs. AAF97787 to AAF97829 represent PCR primers used in

CC the exemplification of the present invention, and AAF97830 to AAF97874

CC represent sequences given in the exemplification of the present

CC invention.

XX Sequence 11967 BP; 2877 A; 2760 C; 2873 G; 3457 T; 0 other;

SQ Query Match 0.2%; Score 55; DB 22; Length 11967;

Best Local Similarity 100.0%; Pred. No. 3.9e-09;

Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17396 ggagtgcgaatggcgtgatctcagctcactgcaacctccgcctcccggttcaagc 17450

DB 8878 ggagtgcgaatggcgtgatctcagctcactgcaacctccgcctcccggttcaagc 8932

RESULT 57

AAT71699

ID AAT71699 standard; DNA; 20303 BP.

XX AAT71699;

XX 20-AUG-1997 (first entry)

XX Human deoxycytidylate deaminase intron 2 encoding DNA.

DE Recombinant deaminase; dCMP; ds.

KW Homo sapiens.

XX US5622851-A.

XX 22-APR-1997.

XX 10-JAN-1995; 95US-0370975.

XX 10-JAN-1995; 95US-0370975.

XX (HEAL-) HEALTH RES INC.

XX Maley F, Maley GR, Weiner KXB;

XX WPI; 1997-244391/22.

XX DNA encoding human deoxycytidylate deaminase - for production of

PT recombinant deaminase

XX Claim 2; Column 83-100; 58pp; English.

XX The present sequence encodes the human deoxycytidylate (dCMP)

CC deaminase intron 2, which comprises 20303 base pairs from nucleotides

CC 1964-22266 of the dCMP deaminase sense strand. The dCMP deaminase gene

CC contains a 5' untranslated region (including the promoter), 5 exons,

CC 4 introns and a 3' untranslated region (including the stop signals).

CC The gene can be used to produce recombinant dCMP deaminase, which can

CC be used to convert dCMP to dUMP. Also, the dCMP gene can be altered

CC (removed or mutated) to alter DNA replication in cells, which may lead

CC to mutagenesis.

XX Sequence 20303 BP; 5454 A; 4115 C; 5052 G; 5682 T; 0 other;

SQ Query Match 0.2%; Score 55; DB 18; Length 20303;

Best Local Similarity 100.0%; Pred. No. 3.5e-09;

Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9013 gttcaatttttttagtagatgggtttccaccatgttggccagctggt 9067


```
Query Match          0.2%; Score 54; DB 21; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttttagtagagatgggtttcaccatgttgccaggctggtt 9068
      |||||||
Db 84  TAAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTGGCCAGCTGGTT 31

RESULT 60
AAC00358
ID AAC00358 standard; cDNA; 354 BP.
AC AAC00358;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 356.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
OS
XX EP1033401-A2.
PN
XX 06-SEP-2000.
PD
XX 21-FEB-2000; 2000EP-0200610.
PF
XX 26-FEB-1999; 99US-0122487.
PR
XX (GEST ) GENSET.
PA
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX WPI: 2000-500381/45.
DR
XX P-PSDB; AAC00352.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 356; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dr primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
XX Sequence 354 BP; 79 A; 94 C; 79 G; 102 T; 0 other;

Query Match          0.2%; Score 54; DB 21; Length 354;
Best Local Similarity 100.0%; Pred. No. 1.6e-08;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17555 caggctgggtcgaactcctgaactcaggtgatccaccacactcagcctccaa 17608
      |||||||
Db 265 caggctgggtcgaactcctgaactcaggtgatccaccacactcagcctccaa 318
```

```
RESULT 61
AAF54060/c
ID AAF54060 standard; DNA; 723 BP.
XX
XX AAF54060;
AC
XX 30-MAR-2001 (first entry)
XX
XX hFIX gene AE3' age-related regulatory region fragment, SEQ ID NO:59.
XX
XX Age-related gene regulation; specific; gene expression;
KW human factor IX; hFIX; AE3'; 3' UTR; 3' untranslated region;
KW AE3'' element; age-regulatable expression construct;
KW antisenese therapy; gene therapy; thrombosis; cardiovascular disease;
KW diabetes; Alzheimer's disease; Parkinson's disease; cancer; osteoporosis;
KW osteoarthritis; dementia; ds.
XX
XX Homo sapiens.
OS
XX WO200075279-A2.
PN
XX 14-DEC-2000.
PD
XX 06-JUN-2000; 2000WO-US15728.
PF
XX 09-JUN-1999; 99US-0328925.
PR
XX (UNMI ) UNIV MICHIGAN.
PA
XX Kurachi K, Kurachi S;
PI
XX WPI: 2001-061708/07.
DR
XX
XX New regulatory elements that control age-related gene expression,
PT useful in gene therapy and for reducing Factor IX expression -
PT
XX Disclosure; Page 177; 225pp; English.
XX
XX The invention relates to nucleic acid sequences which regulate gene
CC expression in an age-related manner and/or in a liver-specific manner.
CC The invention identifies regions of the human factor IX (hFIX) gene, and
CC a region of the human protein C (hPC) gene, which are age-related
CC regulatory sequences. The hFIX age-related regulatory sequences are
CC designated AE5' (AAF54016) and AE3' (AAF54017) and are found in the 5'
CC UTR (at position 2164-2165 of AAF54016) and 3' UTR (at position
CC 34383-35655 of AAF54018) respectively. These elements act synergistically
CC to increase hFIX levels over the lifespan of an individual; however, they
CC can independently exert effects on hFIX mRNA in an age-related manner,
CC with AE5' acting to stabilise hFIX mRNA, and AE3' acting to increase hFIX
CC mRNA levels, over time. AE5' also directs liver-specific expression. The
CC hPC gene age-related regulatory sequence is found in the 5' UTR
CC (AAF54081), and contains two PEA-3 (polyoma virus activator 3) elements
CC 5'-GAGGAAA-3' and 5'-CAGGAAG-3'. The age-related regulatory sequences of
CC the invention, along with their homologues, variants and fragments, may
CC be used in the construction of recombinant expression vectors for the
CC expression of a desired sequence in an age-related fashion in a host
CC cell. Preferred target genes for expression in such age-regulatable
CC expression vectors include those encoding proteins involved in blood
CC coagulation (e.g., the pro-coagulants factor IX and factor VIII, and the
CC anti-coagulants protein C and antithrombin III), human
CC alpha-1-antitrypsin, PEA-3 protein and reporter proteins such as
CC luciferase. Preferred promoters for use in such age-regulatable
CC expression vectors include the human factor IX promoter, the T7 promoter,
CC the T3 promoter and the SP6 promoter. The expression vectors of the
CC invention may be used in gene therapy to provide age- related and/or
CC liver-specific expression of target genes. Age-regulatable constructs may
CC be used in the treatment of such age-related conditions such as
CC thrombosis, cardiovascular disease, diabetes, Alzheimer's disease,
CC Parkinson's disease, cancer, osteoporosis, osteoarthritis and dementia.
CC Specifically, they may be used to express factor IX antisense mRNA in the
CC treatment of thrombotic conditions associated with the natural
CC age-related rise in factor IX expression. Transgenic cells or animals
```

CC that contain vectors of the invention are useful as models of these
CC diseases, in screening for potential therapeutic agents and for studying
CC normal processes such as ageing and gene expression. Fragments and
CC homologues of age-related regulatory sequences, are useful as probes to
CC detect, isolate or identify other such sequences in samples. The present
CC sequence represents an AE3' region fragment.

XX Sequence 723 BP; 213 A; 147 C; 179 G; 184 T; 0 other;

Query Match 0.2%; Score 54; DB 22; Length 723;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9009 ccagtgtaattttgtatttttagtagagatggggtttccacatgttggccagg 9062

Db 258 CCCAGTTAAATTTTGTATTTTAGTAGAGATGGGGTTTCACCATGTGGCCAGG 205

RESULT 62

AAF54018/C

ID AAF54018 standard; DNA; 38059 BP.

XX AAF54018;

XX 30-MAR-2001 (first entry)

XX Human factor IX (hFIX) gene, SEQ ID NO:4.

XX Age-related gene regulation; liver-specific; gene expression;
KW human factor IX; hFIX; AE5'; AE3'; age-regulatable expression construct;
KW antisense therapy; gene therapy; thrombosis; cardiovascular disease;
KW diabetes; Alzheimer's disease; Parkinson's disease; cancer; osteoporosis;
KW osteoarthritis; dementia; ds.

XX Homo sapiens.

XX WO200075279-A2.

XX 14-DEC-2000.

XX 06-JUN-2000; 2000WO-US15728.

XX 09-JUN-1999; 99US-0328925.

XX (UNMI) UNIV MICHIGAN.

XX Kurachi K, Kurachi S;

XX WPI; 2001-061708/07.

DR P-PSDB: AAB60281, AAB60282, AAB60283, AAB60284, AAB60285, AAB60286,

DR AAB60287, AAB60288, AAB60289.

XX New regulatory elements that control age-related gene expression,
PT useful in gene therapy and for reducing Factor IX expression -

XX Disclosure; Fig 8A-E; 225pp; English.

XX The invention relates to nucleic acid sequences which regulate gene
CC expression in an age-related manner and/or in a liver-specific manner.
CC The invention identifies regions of the human factor IX (hFIX) gene, and
CC a region of the human protein C (hpc) gene, which are age-related
CC regulatory sequences. The hFIX age-related regulatory sequences are
CC designated AE5' (AAF54016) and AE3' (AAF54017) and are found in the 5'
CC UTR (at position 2164-2165 of AAF54018) and 3' UTR (at position
CC 34383-35655 of AAF54018) respectively. These elements act synergistically
CC to increase hFIX levels over the lifespan of an individual; however, they
CC can independently exert effects on hFIX mRNA in an age-related manner,
CC with AE5' acting to stabilise hFIX mRNA, and AE3' acting to increase hFIX
CC mRNA levels, over time. AE5' also directs liver-specific expression. The
CC hpc gene age-related regulatory sequence is found in the 5' UTR
CC (AAF54081), and contains two PEA-3 (polyoma virus activator 3) elements
CC 5'-GAGGAAA-3' and 5'-CAGGAAG-3'. The age-related regulatory sequences of

CC the invention, along with their homologues, variants and fragments, may
CC be used in the construction of recombinant expression vectors for the
CC expression of a desired sequence in an age-related fashion in a host
CC cell. Preferred target genes for expression in such age-regulatable
CC expression vectors include those encoding proteins involved in blood
CC coagulation (e.g., the pro-coagulants factor IX and factor VIII, and the
CC anti-coagulants protein C and antithrombin III), human
CC alpha-1-antitrypsin, PEA-3 protein and reporter proteins such as
CC luciferase. Preferred promoters for use in such age-regulatable
CC expression vectors include the human factor IX promoter, the T7 promoter,
CC the T3 promoter and the SP6 promoter. The expression vectors of the
CC invention may be used in gene therapy to provide age-related and/or
CC liver-specific expression of target genes. Age-regulatable constructs may
CC be used in the treatment of such age-related conditions such as
CC thrombosis, cardiovascular disease, diabetes, Alzheimer's disease,
CC Parkinson's disease, cancer, osteoporosis, osteoarthritis and dementia.
CC Specifically, they may be used to express factor IX antisense mRNA in the
CC treatment of thrombotic conditions associated with the natural
CC age-related rise in factor IX expression. Transgenic cells or animals
CC that contain vectors of the invention are useful as models of these
CC diseases, in screening for potential therapeutic agents and for studying
CC normal processes such as ageing and gene expression. Fragments and
CC homologues of age-related regulatory sequences, are useful as probes to
CC detect, isolate or identify other such sequences in samples. The present
CC sequence represents the hFIX gene.

XX Sequence 38059 BP; 12326 A; 7397 C; 7441 G; 10895 T; 0 other;

Query Match 0.2%; Score 54; DB 22; Length 38059;
Best Local Similarity 100.0%; Pred. No. 6.5e-09;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9009 ccagtgtaattttgtatttttagtagagatggggtttccacatgttggccagg 9062

Db 31675 CCCAGTTAAATTTTGTATTTTAGTAGAGATGGGGTTTCACCATGTGGCCAGG 31622

RESULT 63

AAT25953

ID AAT25953 standard; cDNA to mRNA; 72 BP.

XX AAT25953;

XX 28-OCT-1996 (first entry)

XX Human gene signature HUMGS08188.

XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.

XX Homo sapiens.

XX WO9514772-A1.

XX 01-JUN-1995.

XX 11-NOV-1994; 94WO-JP01916.

XX 12-NOV-1993; 93JP-0355504.

XX (MATS/) MATSUBARA K.

XX (OKUB/) OKUBO K.

XX Matsubara K, Okubo K;

XX WPI; 1995-206931/27.

XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues

XX Claim 1; Page 1967; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in AAT19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

XX SQ Sequence 72 BP; 25 A; 10 C; 16 G; 20 T; 1 other;

Query Match 0.2%; Score 53; DB 16; Length 72;
Best Local Similarity 100.0%; Pred. No. 4.8e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26414 tcgagacttaacgaaatagatttcagctgcaataaagattgattgcaa 26466
|||||

Db 19 tcgagacttaacgaaatagatttcagctgcaataaagattgattgcaa 71

RESULT 64

AAC30074

ID AAC30074 standard; cDNA; 192 BP.

XX AC AAC30074;

XX 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 34149.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
gene therapy; chromosome mapping; ss.

KW Homo sapiens.

OS Homo sapiens.

XX EP1033401-A2.

PN 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

PI WPI; 2000-500381/45.

DR New nucleic acid that is a 5' expressed sequence tag (5' EST) for

XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 34149; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. No ORF has yet been conclusively

CC identified within the present sequence. The 5' ESTs were prepared from

CC total human RNAs or poly(A) RNAs derived from 30 different tissues. EST

CC sequences usually correspond mainly to the 3' untranslated region (UTR)

CC of the mRNA because they are often obtained from oligo-dT primed cDNA

CC libraries. Such ESTs are not well suited for isolating cDNA sequences

CC derived from the 5' ends of mRNAs and even in those cases where longer

CC cDNA sequences have been obtained, the full 5' UTR is rarely included.

CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.

CC They are used to obtain upstream regulatory sequences and to design

XX expression and secretion vectors.

XX SQ Sequence 192 BP; 31 A; 48 C; 45 G; 65 T; 3 other;

Query Match 0.2%; Score 53; DB 21; Length 192;
Best Local Similarity 100.0%; Pred. No. 3.9e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtatttttagtagagatgggtttccaccatgttggccaggctggt 9067
|||||

Db 37 taattttgtatttttagtagagatgggtttccaccatgttggccaggctggt 89

RESULT 65

AZ13926

ID AZ13926 standard; cDNA; 300 BP.

XX AC AZ13926;

XX 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:1395.

XX Human; gene; gene expression product; diagnosis; therapy; probe;
detection; mapping; tissue typing; profiling; forensic; cancer;
genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

KW Homo sapiens.

XX WO9938972-A2.

PN 05-AUG-1999.

PD 28-JAN-1999; 99WO-US01619.

XX 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

DR Novel human genes and their expression products which are

XX differentially expressed in different cell types

PT Claim 1; Page 977; 2479pp; English.

XX The present invention describes a library of human polynucleotides

CC comprising the sequences given in AA212532 to AA217779. Also described is

CC a method of detecting differentially expressed genes correlated with the

CC cancerous state of a mammalian cell, comprising detecting at least one

CC differentially expressed gene product in a test sample from a cell

CC suspected of being cancerous, where the gene product is encoded by one

CC of the 5248 polynucleotide sequences given in AA212532 to AA217779. The

CC polynucleotides can be used as a source of primers and probes, which can

CC be used for a variety of purpose, e.g. detection of expression levels,

CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.

XX
SQ Sequence 300 BP; 79 A; 72 C; 81 G; 68 T; 0 other;

Query Match 0.2%; Score 53; DB 20; Length 300;
Best Local Similarity 100.0%; Pred. No. 3.6e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17560 tggctgaactcctgacctcaggtgattccaccacccacctcagctcccaagtg 17612

Db 80 tggctgaactcctgacctcaggtgattccaccacccacctcagctcccaagtg 132

RESULT 66
AAF65526/c
ID AAF65526 standard; cDNA; 350 BP.

XX AAF65526;

XX 09-APR-2001 (first entry)

XX Novel human polynucleotide, SEQ ID NO: 1282.

XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

XX WO200102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000WO-US18374.

XX 02-JUL-1999; 99US-0142310.

XX 02-JUL-1999; 99US-0142311.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI Kita D, Garcia V, Jones LW, Strache-Crain B;

XX WPI; 2001-091805/10.

XX Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -

XX Claim 9; Page 727; 1046pp; English.

XX The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and

CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.

XX Sequence 350 BP; 103 A; 71 C; 98 G; 77 T; 1 other;

Query Match 0.2%; Score 53; DB 22; Length 350;
Best Local Similarity 100.0%; Pred. No. 3.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagtagatgggggtttccaccatgttggcagctggt 9067

Db 265 TAATTTTGTATTTTAGTAGATGGGGTTCACCATGTTGGCCAGCTGCT 213

RESULT 67
AAF65527/c
ID AAF65527 standard; cDNA; 352 BP.

XX AAF65527;

XX 09-APR-2001 (first entry)

XX Novel human polynucleotide, SEQ ID NO: 1283.

XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

XX WO200102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000WO-US18374.

XX 02-JUL-1999; 99US-0142310.

XX 02-JUL-1999; 99US-0142311.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI Kita D, Garcia V, Jones LW, Strache-Crain B;

XX WPI; 2001-091805/10.

XX Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -

XX Claim 9; Page 727; 1046pp; English.

XX The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to

CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
SQ Sequence 352 BP; 103 A; 72 C; 98 G; 79 T; 0 other;

Query Match 0.2%; Score 53; DB 22; Length 352;
Best Local Similarity 100.0%; Pred. No. 3.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtatttagtagagatggggtttccaccatgttggccaggtggt 9067
|||||
Db 265 TAATTTTGTATTTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGT 213

RESULT 68
AAX79050/c
ID AAX79050 standard; DNA; 541 BP.
XX
AC AAX79050;
XX
DT 17-AUG-1999 (first entry)
XX
DE Human secreted protein gene 40 clone HFIUR10.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; ischaemia; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
XX WO9919339-A1.
PN
XX
PD 22-APR-1999.
XX
XX
PF 08-OCT-1998; 98WO-US21142.
XX
XX 09-OCT-1997; 97US-0071498.
PR
XX 09-OCT-1997; 97US-0061463.
PR
XX 09-OCT-1997; 97US-0061527.
PR
XX 09-OCT-1997; 97US-0061529.
PR
XX 09-OCT-1997; 97US-0061532.
PR
XX 09-OCT-1997; 97US-0061536.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Brewer LA, Duan R, Ebner R, Ferrie AM, Florence C;
PI Florence KA, Greene JM, Olsen HS, Rosen CA, Ruben SM;
PI Young PE, Yu G;
XX
XX WPI: 1999-277587/23.
DR
XX P-PSDB; AAY14450.
DR
XX
XX
XX New isolated human genes and the secreted polypeptides they encode
PT
XX
XX Claim 1; Page 177; 226pp; English.
PS
XX
XX This sequence represents a nucleic acid molecule which encodes a
CC secreted human protein. The gene number, and the clone it is derived
CC from, are detailed in the descriptor line. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. AAX79002) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 53 novel genes and their fragments (nucleic
CC acid sequences: AAX79011-X79064; amino acid sequences AAY14411-Y14464)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new

CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 53
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAX79011 for described uses).
XX
SQ Sequence 541 BP; 109 A; 170 C; 149 G; 113 T; 0 other;

Query Match 0.2%; Score 53; DB 20; Length 541;
Best Local Similarity 100.0%; Pred. No. 3.2e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtatttagtagagatggggtttccaccatgttggccaggtggt 9067
|||||
Db 484 TAATTTTGTATTTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGT 432

RESULT 69
AAA44206
ID AAA44206 standard; CDNA; 651 BP.
XX
AC AAA44206;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:781.
XX
KW Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
KW antilucer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell osteoarthritis;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200021991-A1.
PN
XX
XX 20-APR-2000.
PD
XX
XX 15-OCT-1999; 99WO-US24206.
PF
XX
XX 15-OCT-1998; 98US-0104436.
PR
XX
XX (GEMY) GENETICS INST INC.
PA
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR;
PI
XX WPI: 2000-317938/27.
DR
XX
XX Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (sESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
XX Claim 1; Page 407; 803pp; English.
PS
XX
XX AAA44206 to AAA45925 represent specifically claimed secreted expressed
CC sequence tags (sESTs), isolated from human, mouse, chicken and rat
CC tissue sources. The sESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytostatic; antibacterial; antidiabetic; antipsoriatic; antidiabetic;
CC antiasthmatic; vulnery; antilucer; osteopathic; neuroprotective;

CC nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The SESTs can be used for gene
CC therapy and in vaccines. The SESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention.

XX Sequence 651 BP; 141 A; 150 C; 140 G; 219 T; 1 other;

Query Match 0.2%; Score 53; DB 21; Length 651;
Best Local Similarity 100.0%; Pred. No. 3.1e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagatagatggggtttccaccatgttgccagctggt 9067
|||||
Db 215 taattttgtatttttagatagatggggtttccaccatgttgccagctggt 267

RESULT 70
AAAX99010/c

ID AAAX99010 standard; cDNA; 746 BP.

AC AAAX99010;

DT 24-SEP-1999 (first entry)

DE Human validated cancer cell derived CDNA #332.

KW Cancer; human; colon; breast; lung; transmembrane receptor; ATPase;
KW integral membrane protein; aspartyl protease; GATA family; wnt family;
KW transcription factor; G-protein alpha subunit; protein phosphatase;
KW phorbolster binding protein; diacylglycerol binding protein; trypsin;
KW protein kinase; tyrosine phosphatase; developmental signalling protein;
KW WW/rsp5/WWP domain; therapy; forensic; genetic mapping; diagnostic;
KW detection; treatment; cervical; melanoma; colorectal adenocarcinoma;
KW Wilm's tumour; retinoblastoma; sarcoma; myosarcoma; lung carcinoma;
KW leukemia; lymphoma; dysplasia; hyperplasia; endometrium; adrenal;
KW prostate; ss.

XX Homo sapiens.

XX WO9933982-A2.

XX 08-JUL-1999.

XX 22-DEC-1998; 98WO-US27610.

XX 21-DEC-1998; 98US-0217471.

XX 23-DEC-1997; 97US-0068755.

XX 03-APR-1998; 98US-0080664.

XX 21-OCT-1998; 98US-0105234.

XX 27-OCT-1998; 98US-0105877.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

XX Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

XX Jones LW, Kassam A, Kennedy GC, Kita D, Labat I;

XX Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

XX Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-430243/36.

XX New isolated human polynucleotides
PT
XX
XX Claim 1; Page 546; 591pp; English.

CC This invention describes novel isolated human polynucleotides obtained
CC by screening for differential expression in colon cancer, breast cancer
CC and lung cancer cell lines. The polynucleotides of the invention are
CC represented in AAX98275-X99118 and encode polypeptides of protein
CC families selected from 4 transmembrane segments integral membrane
CC proteins, 7 transmembrane receptors, ATPases associated with various
CC cellular activities (AAA), eukaryotic aspartyl proteases, GATA family of
CC transcription factors, G-protein alpha subunit, phorbolsters or
CC diacylglycerol binding proteins, protein kinase, protein phosphatase 2C,
CC protein tyrosine phosphatase, trypsin, wnt family of developmental
CC signalling proteins and WW/rsp5/WWP domain containing proteins. The
CC encoded polypeptides also have a functional domain selected from Ank
CC repeat, basic region plus leucine zipper transcription factors,
CC bromodomain, EF-hand, SH3 domain, WD domain/G-beta repeats, zinc finger
CC (C2H2 type), zinc finger (CCHC class), and zinc-binding metalloprotease
CC domain. The polynucleotides encode polypeptides with similarity to known
CC protein families and are predicted to have similar properties. The novel
CC polynucleotides can be used to develop products for use as therapeutic
CC agents and in forensics, genetic analysis, mapping and diagnostic
CC applications. In particular, the product can be used for the detection
CC and management of cancers. They can be used for treating e.g. cervical
CC cancers, melanomas, colorectal adenocarcinomas, Wilm's tumour, sarcomas,
CC retinoblastoma, myosarcomas, lung carcinomas, leukemias, such as chronic
CC myelogenous leukemia, promyelocytic leukemia, monocytic leukemia, and
CC myeloid leukemia, and lymphomas such as histiocytic lymphoma, anhydric
CC hereditary ectodermal dysplasia, congenital alveolar dysplasia,
CC epithelial dysplasia of the cervix, fibrous dysplasia of bone, and
CC mammary dysplasia, hyperplasias, e.g. endometrial, adrenal, breast,
CC prostate or thyroid hyperplasias or pseudoepitheliomatous hyperplasia of
CC the skin.

XX Sequence 746 BP; 200 A; 151 C; 185 G; 185 T; 25 other;

Query Match 0.2%; Score 53; DB 20; Length 746;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagatagatggggtttccaccatgttgccagctggt 9067
|||||

Db 470 TAATTTTGTATTTTATTAGATGGGGTTTCACCATGTGGCCAGCTGGT 418

RESULT 71

AAC74381/c

ID AAC74381 standard; cDNA; 788 BP.

XX AAC74381;

XX 02-FEB-2001 (first entry)

XX Human secreted protein gene 45 SEQ ID NO:55.

DE Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; gene therapy;
KW food additive; preservative; chromosome identification; cancer;
KW female reproductive system disorder; immune disorder; wound healing;
KW cardiovascular disorder; neurological disease; infectious disease;
KW infection; ss.

XX Homo sapiens.

XX WO200058340-A2.

XX 05-OCT-2000.

PF 23-MAR-2000; 2000WO-US07724.
XX
PR 26-MAR-1999; 99US-0126510.
PR 07-JAN-2000; 2000US-0174850.
XX
PR
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI: 2000-594638/56.
DR P-PSDB; AAB39446.
XX
XX Fifty nucleic acid molecules encoding human secreted proteins, useful
PT in the prevention, treatment and diagnosis of cancer, immune disorders,
PT cardiovascular disorders and neurological diseases.
XX
XX Claim 1; Page 353; 391pp; English.
XX
XX The polynucleotide sequences given in AAC74337 to AAC74386 encode the
CC human secreted proteins given in AAB39402 to AAB39451. AAB39452 to
CC AAB39484 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Example of activities include:
CC cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
CC antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
CC vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic; and
CC cardiant. The polynucleotides and polypeptides are useful for preventing,
CC treating or ameliorating a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. The polypeptides
CC can also be used as a food additive or preservative to increase or
CC decrease storage capabilities. The polynucleotide are useful for
CC chromosome identification. They are also useful as probes for diagnosing
CC a disorder related to the female reproductive system, particularly breast
CC and/or ovary cancer. They are also useful in the gene therapy of breast
CC and ovarian cancer. Secreted protein nucleic acids, proteins,
CC antibodies, agonists and antagonists are useful in the diagnosis,
CC treatment and prevention of: (a) cancer; (b) immune disorders; (c)
CC cardiovascular disorders; (d) wound healing; (e) neurological diseases;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections. AAC74328 to AAC74336 and AAB39401 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 788 BP; 196 A; 174 C; 224 G; 194 T; 0 other;

Query Match 0.2%; Score 53; DB 21; Length 788;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taatttttatttttagtagagatgggtttccaccatgttggccaggctggt 9067
|||||
Db 624 TAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGT 572

RESULT 72
AAC81263/c
ID AAC81263 standard; DNA; 1024 BP.
XX
AC AAC81263;
XX
XX
DT 23-FEB-2001 (first entry)
XX
DE Human tyrosine phosphatase HD-PTP gene exon 2, SEQ ID NO:41.
XX
XX Human; histidine domain-protein tyrosine phosphatase; HD-PTP;
KW chromosome 3p21.3; gene deletion; tumour suppressor; cytostatic;
KW lung cancer; tumour; gene therapy; diagnosis; recombinant production;
KW anticancer; ds.
XX
OS Homo sapiens.
XX
XX WO200063392-A1.

XX 26-OCT-2000.
PD
XX
XX 14-APR-2000; 2000WO-JP02455.
PF
XX
XX 16-APR-1999; 99JP-0108842.
PR
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Shimizu K;
PI
XX
XX WPI: 2000-672740/65.
DR P-PSDB; AAB29666.
XX
XX Human tyrosine phosphatase with oncogenic activity encoded by a gene
PT frequently deleted in lung cancer, useful for treatment and diagnosis
PT of tumors.
XX
XX Claim 4; Page 129-130; 134pp; Japanese.
PS
XX The invention relates to a novel human tyrosine phosphatase, histidine
CC domain-protein tyrosine phosphatase (HD-PTP; AAB29666) and to human
CC HD-PTP nucleic acids (AAC81224, AAC81225, AAC81262, AAC81263). The
CC HD-PTP gene is located on chromosome 3p21.3. This region is frequently
CC found to be deleted in lung cancers, and is therefore thought to contain
CC a tumour suppressor gene. The invention also relates to expression
CC vectors and host cells containing human HD-PTP nucleic acids; the
CC recombinant production of HD-PTP; anticancer drugs containing HD-PTP;
CC gene therapy compositions containing DNA encoding HD-PTP; diagnostic
CC reagents containing HD-PTP oligonucleotides; antibodies specific for
CC HD-PTP; and an immunoassay method using HD-PTP-specific antibodies for
CC use in cancer diagnosis and investigation. HD-PTP proteins, nucleic acids
CC and antibodies may be used in the treatment, investigation and diagnosis
CC of cancers, particularly those of the lung. The present sequence
CC represents exon 2 of the human HD-PTP gene.
XX
SQ Sequence 1024 BP; 241 A; 234 C; 287 G; 262 T; 0 other;

Query Match 0.2%; Score 53; DB 21; Length 1024;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taatttttatttttagtagagatgggtttccaccatgttggccaggctggt 9067
|||||
Db 689 TAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGT 637

RESULT 73
AAA96485/c
ID AAA96485 standard; cDNA; 1721 BP.
XX
XX AAA96485;
AC
XX
XX 08-FEB-2001 (first entry)
DT
XX
DE cDNA encoding a human transmembrane protein.
XX
XX Human; transmembrane protein; cell proliferation disorder; myeloma;
KW reproductive disorder; smooth muscle disorder; neurological disorder;
KW arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
KW allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KW Alzheimer's disease; Tourette's disorder; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 276..1151
FT /*tag= a
FT 276..440
FT sig_peptide /*tag= b
XX
XX WO200056891-A2.

XX PD 28-SEP-2000.
XX PF
XX PR 22-MAR-2000; 2000WO-US07817.
XX PR 22-MAR-1999; 99US-0125537.
XX PR 16-JUN-1999; 99US-0139565.
XX PA (INCY-) INCYTE PHARM INC.
XX PI Yue H, Lal P, Tang YT, Hillman JL, Reddy R, Bandman O, Baughn MR;
XX PI Lu DAM, Azimzai Y, Yang J;
XX DR WPI: 2000-579485/54.
XX DR P-PSDB; AAB18972.
XX PT New human transmembrane proteins are used to treat a disease or
XX PT condition associated with decreased expression of functional HTMP e.g.
XX PT Tourette's disorder, angina and leukaemia -
XX PS Claim 4; Page 116; 130pp; English.
XX CC The present sequence encodes a human transmembrane proteins (HTMP).
XX CC Agonists and antagonists of the protein are used to treat a disease
XX CC or condition associated with overexpression of the protein. Diseases
XX CC and conditions which can be treated include cell proliferative,
XX CC immunological, reproductive, smooth muscle and neurological disorders
XX CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
XX CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
XX CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
XX CC polynucleotides may be used to detect and quantify gene expression in
XX CC biopsied tissues where protein expression may be correlated with disease
XX CC e.g. to determine absence, presence or excess expression of HTMP or to
XX CC monitor regulation of HTMP expression during therapeutic intervention.
XX SQ Sequence 1721 BP; 306 A; 579 C; 503 G; 333 T; 0 other;

Query Match 0.2%; Score 53; DB 21; Length 1721;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagtagagatgggtttccaccatgttgccagctggt 9067
|||||
Db 1617 TAAATTTGTATTTTGTAGATGGGGTTTCACCATGTTGGCCAGCTGTT 1565
|||||

RESULT 74
AAF93811
ID AAF93811 standard; cDNA; 1773 BP.
XX AC AAF93811;
XX AC
XX DT 23-MAY-2001 (first entry)
XX DE
XX DE Human cDNA encoding a membrane or secretory protein clone PSEC0127.
XX KW Human; secretory protein; membrane protein; vaccine; gene therapy;
XX KW rheumatoid arthritis; diabetes; ss.
XX OS Homo sapiens.
XX PN EP1067182-A2.
XX XX
XX PD 10-JAN-2001.
XX PF 07-JUL-2000; 2000EP-0114090.
XX PR 08-JUL-1999; 99JP-0194179.
XX PR 11-JAN-2000; 2000JP-0118775.
XX PR 02-MAY-2000; 2000JP-0183766.
XX PA (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX WPI: 2001-093989/11.
XX DR P-PSDB; AAB88384.
XX XX
XX PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
XX PT gene therapy or as candidate target molecules in drug development -
XX PS Claim 1; SEQ ID 135; 609pp + CD ROM; English.
XX XX
XX CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
XX CC which encode human secretory or membrane proteins represented by
XX CC AAB88317 - AAB88419. Included in the invention are primers
XX CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
XX CC cDNA sequences of the invention. The invention also includes methods for
XX CC the production of antibodies directed against the proteins, and cDNA
XX CC sequences, which can be used in vaccines. The polynucleotide sequences
XX CC can be used in gene therapy. The polynucleotide sequences and the
XX CC proteins they encode may be used in the prevention, treatment and
XX CC diagnosis of diseases associated with inappropriate secretory
XX CC protein/membrane protein expression. The nucleic acids and complementary
XX CC sequences may also be used as DNA probes in diagnostic assays
XX CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
XX CC presence of similar nucleic acid sequences in samples. They may also be
XX CC used to study the expression and function of secretory proteins/membrane
XX CC polypeptides and their role in metabolism. The polypeptides may be used
XX CC as antigens in the production of antibodies against them and in assays to
XX CC identify modulators (agonists and antagonists) of expression and
XX CC activity. The antibodies and antagonists may also be used as therapeutic
XX CC agents to down regulate expression and activity. The antibodies may also
XX CC be used as diagnostic agents for detecting the presence of the
XX CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
XX CC (ELISA). Examples of diseases which may be treated include rheumatoid
XX CC arthritis and diabetes.
XX SQ Sequence 1773 BP; 489 A; 384 C; 417 G; 483 T; 0 other;

Query Match 0.2%; Score 53; DB 22; Length 1773;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagtagagatgggtttccaccatgttgccagctggt 9067
|||||
Db 918 taattttgtatttttagtagagatgggtttccaccatgttgccagctggt 970
|||||

RESULT 75
AAC88096/C
ID AAC88096 standard; cDNA; 1991 BP.
XX AC AAC88096;
XX AC
XX DT 09-MAR-2001 (first entry)
XX DE
XX DE Human FLEXHT-27 nucleotide sequence SEQ ID NO:82.
XX KW Human; FLEXHT; full-length molecules expressed in human tissue;
XX KW diagnosis; gene expression; genetic linkage; genetic variability;
XX KW antianaemic; anticonvulsant; antiarteriosclerotic; immunomodulatory;
XX KW cytostatic; antiasthmatic; antiinflammatory; hepatotropic; antidiabetic;
XX KW anti-gout; antithyroid; neuroprotective; antiarthritic; osteopathic;
XX KW antipsoriatic; antirheumatic; antiulcer; gene therapy; anaemia; gout;
XX KW epilepsy; arteriosclerosis; atherosclerosis; developmental disorder;
XX KW cancer; immunological disorder; asthma; bronchitis; cirrhosis;
XX KW Crohn's disease; diabetes mellitus; Grave's disease; multiple sclerosis;
XX KW osteoarthritis; pancreatitis; rheumatoid arthritis; psoriasis;
XX KW ulcerative colitis; ss.
XX OS Homo sapiens.
XX PN WO200070047-A2.


```
XX 19-MAR-1998; 98US-0078563.
PR 19-MAR-1998; 98US-0078566.
PR 19-MAR-1998; 98US-0078573.
PR 19-MAR-1998; 98US-0078574.
PR 19-MAR-1998; 98US-0078576.
PR 19-MAR-1998; 98US-0078577.
PR 19-MAR-1998; 98US-0078578.
PR 19-MAR-1998; 98US-0078579.
PR 19-MAR-1998; 98US-0078581.
PR 01-APR-1998; 98US-0080312.
PR 01-APR-1998; 98US-0080313.
PR 01-APR-1998; 98US-0080314.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;
PI Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW;
PI Olsen HS, Shi Y, Moore PA;
XX
XX WPI; 1999-562050/47.
DR P-PSDB; AAY41396.
XX
XX New isolated human genes, useful for diagnosis and treatment of e.g.
PT cancers, neurological disorders, immune diseases, inflammation or blood
PT disorders -
XX
XX Claim 1; Page 350-351; 484pp; English.
PS
XX
XX This sequence represents a nucleic acid molecule which encodes a
CC secreted human protein. The gene number, and the clone it is derived
CC from, are detailed in the descriptor line. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. AA224802) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC
CC The invention relates to 95 novel genes and their fragments (nucleic
CC acid sequences: AA224811-224907; amino acid sequences AAY41308-Y41404)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 95
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AA224811 for described uses).
XX
XX Sequence 2596 BP; 790 A; 433 C; 500 G; 873 T; 0 other;
SQ
Query Match 0.2%; Score 53; DB 20; Length 2596;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9015 taattttgtatttttagtagagatggggtttaccatgttgccagcgctggt 9067
Db 1573 taattttgtatttttagtagagatggggtttaccatgttgccagcgctggt 1625
RESULT 78
AAF21203/c
ID AAF21203 standard; DNA; 5996 BP.
XX
XX AAF21203;
AC
XX
XX 14-MAR-2001 (first entry)
DT
XX
XX Human low adenosine antisense oligonucleotide related sequence #2770.
DE
XX
XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary vasoconstriction; impeded respiration;
KW surfactant hypo-production; pulmonary vasoconstriction; asthma; RDS;
```

```
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
XX Homo sapiens.
OS
XX WO2000062736-A2.
XX
XX 26-OCT-2000.
PD
XX
XX 24-MAR-2000; 2000WO-US08020.
XX
XX 06-APR-1999; 99US-0127958.
XX
XX (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
XX Nyce JW;
XX
XX WPI; 2000-679539/66.
DR
XX
XX Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
XX Disclosure; Page 1117-1118; 1592pp; English.
PS
XX
XX The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypo-production which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
XX Sequence 5996 BP; 1756 A; 1275 C; 1566 G; 1399 T; 0 other;
SQ
Query Match 0.2%; Score 53; DB 21; Length 5996;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9015 taattttgtatttttagtagagatggggtttaccatgttgccagcgctggt 9067
Db 1138 TAATTTTGTATTTTTAGTAGATGGGTTTCCACCATGTTGGCAGGCTGGT 1086
RESULT 79
AAA35081/c
ID AAA35081 standard; DNA; 6056 BP.
```


KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; BAC; transgenic;
KW human artificial chromosome; bacterial artificial chromosome; ss.
XX Synthetic.
XX WO9851790-A1.
XX 19-NOV-1998.
XX 13-MAY-1998; 98WO-AU003352.
XX 26-AUG-1997; 97AU-0008791.
XX 13-MAY-1997; 97AU-0006784.
XX (AMRA-) AMRAD OPERATIONS PTY LTD.
XX Cancilla MR, Choo K, Du Sart D;
XX WPI; 1999-009773/01.
XX New isolated nucleic acid comprising neocentromere sequences from
XX eukaryotic chromosome - used to produce replicable, segregating
XX artificial chromosomes that can carry large amounts of DNA for gene
XX therapy
XX
XX Claim 10; Page 195-203; 540pp; English.
XX
XX The present sequence represents a bacterial artificial chromosome (BAC)
XX contig, and exemplifies the invention. The specification describes
XX nucleic acid sequences derived from a eukaryotic chromosome, including a
XX neocentromere or its functional derivative or hybrid, that are able, in
XX a compatible cell, of replicating, acting as extra-chromosomal element
XX and segregating during cell division. The sequences can be used to
XX construct artificial chromosomes for use in gene therapy comprising a
XX replicable, segregating nucleic acid that confers a specific phenotype
XX on cells. Human artificial chromosomes can propagate in human cells and
XX carry large amounts of DNA (e.g. therapeutic genes), and, being
XX extra-chromosomal, they are not mutagenic. The artificial chromosomes
XX are also useful for generation of transgenic plants and animals, in
XX production of proteins and to make diagnostic reagents, e.g. for
XX expression of cytokines, receptors and growth factors, or to increase
XX the copy number of a gene in a cell. The constructs may also be
XX used for functional and structural analysis of chromosomes.
XX
XX Sequence 11811 BP; 3014 A; 2459 C; 2433 G; 3905 T; 0 other;
SQ
Query Match 0.2%; Score 53; DB 20; Length 11811;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9015 taattttgtatttttagtagagatgggggtttccaccatgttgccaggtcgtt 9067
|||||
Db 10826 TAAATTTTGTATTTTACTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGT 10774
RESULT 82
AAV40401/C
ID AAV40401 standard; DNA; 13865 BP.
XX
XX AAV40401;
XX
XX 16-OCT-1998 (first entry)
XX
XX Human tissue factor full length genomic DNA sequence.
XX
XX Human; truncated; tissue factor; tTF; TF; tumour; coagulation;
KW blood vessel; Factor VIIa; FVIIa; benign growth; vascularised;
KW benign prostatic hypertrophy; malignant; necrosis; angiogenesis;
KW diabetic retinopathy; restenosis; neovascular glaucoma; psoriasis;
KW rheumatoid arthritis; ss.
XX

OS Homo sapiens.
XX
XX WO9831394-A2.
XX
XX 23-JUL-1998.
XX
XX 20-JAN-1998; 98WO-US01012.
XX
XX 27-MAR-1997; 97US-0042427.
XX 22-JAN-1997; 97US-0035920.
XX 27-JAN-1997; 97US-0036205.
XX
XX (TEXA) UNIV TEXAS SYSTEM.
XX
XX Gao B, King SW, Thorpe PE;
XX
XX WPI; 1998-413821/35.
XX P-PSDB; AAW69613.
XX
XX Composition containing coagulation-defective tissue factor for
XX treating, e.g. tumours - useful for, e.g. promoting coagulation in
XX pro:thrombotic and tumour-associated vasculature, used with, e.g.
XX factor 7 or anti-cancer agent
XX
XX Disclosure; Page 185-193; 225pp; English.
XX
XX A composition has been developed which comprises at least 1 coagulation-
XX deficient tissue factor (TF) compound that is modified to increase its
XX biological half-life, but excluding modification that involves attachment
XX to an antibody (or its antigen-binding region) that binds to a component
XX (cells, vasculature or stroma) of tumours. Also described in the present
XX invention are compositions containing any coagulation-deficient TF for
XX promoting coagulation. The coagulation-deficient TFs are used to promote
XX coagulation preferentially in prothrombotic vessels, particularly those
XX associated with: (i) benign growths (e.g. benign prostatic hypertrophy);
XX (ii) vascularised, malignant tumours of medium or large size (where they
XX also induce tumour necrosis), or (iii) other disorders that involve
XX angiogenesis, e.g. diabetic retinopathy, restenosis, neovascular
XX glaucoma, psoriasis and rheumatoid arthritis. The composition can be
XX administered systemically, particularly intravenously, typically at
XX 0.2-200 mg, given 3 times over 7 days. Truncated TF, and its variants,
XX localise specifically in tumour-associated blood vessels after systemic
XX administration, even though they contain no targeting agent. They cause
XX little if any injury to normal tissue; may produce a synergistic response
XX when used with other antitumour agents and they eliminate the multi-step,
XX and expensive, preparation of antibody-based targeting constructs. The
XX present sequence encodes human TF, from the present invention.
XX
XX Sequence 13865 BP; 3711 A; 2955 C; 3240 G; 3959 T; 0 other;
SQ
Query Match 0.2%; Score 53; DB 19; Length 13865;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9015 taattttgtatttttagtagagatgggggtttccaccatgttgccaggtcgtt 9067
|||||
Db 8526 TAAATTTTGTATTTTACTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGT 8474
RESULT 83
AAZ32165/C
ID AAZ32165 standard; cDNA; 13865 BP.
XX
XX AAZ32165;
XX
XX 13-JAN-2000 (first entry)
XX
XX Human cholesteryl ester transfer nucleotide sequence.
XX
XX Human; coding sequence polymorphism; vascular pathology gene;
KW polymorphic site; phenotype correlation; forensic; paternity testing;
KW medicine; genetic analysis; vascular disease; ds.
KW


```

XX Homo sapiens.
OS
PN WO9950454-A2.
XX
XX PD 07-OCT-1999.
XX
XX PF 26-MAR-1999; 99WO-US06473.
XX
XX PR 01-APR-1998; 98US-0054272.
XX
XX PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
XX PI Lander ES, Daley GQ, Cargill M, Ireland JS, Rozen SG;
XX
XX DR WPI: 1999-620066/53.
XX
XX DR P-PSDB; AAY49556.
XX
XX
XX Determination of polymorphisms in genes, especially those identifying
XX predisposition to vascular disease
XX
XX PS Claim 1; Fig 9; 134pp; English.
XX
XX CC AAZ32159 to AAZ32194 represent reference alleles for specifically
XX claimed nucleic acid sequences from the present invention which comprise
XX polymorphic sites as given in a table in the specification, selected
XX from 92 single nucleotide polymorphisms in which the nucleotide at the
XX polymorphic site is different from a nucleotide at the same site in a
XX reference allele. The nucleic acids, and primers and probes, are used to
XX identify polymorphisms, which may predispose an individual to disease,
XX especially a vascular disease. They can also be used in phenotype
XX correlations, forensics, paternity testing, medicine or genetic
XX analysis. AAY49550 to AAY49573 represent the proteins which correspond
XX to some of the reference alleles.
XX
XX SQ Sequence 13865 BP; 3711 A; 2955 C; 3240 G; 3959 T; 0 other;

Query Match 0.2%; Score 53; DB 20; Length 13865;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtatttagtagagatgggtttccaccatgttgccaggtggt 9067
|||||
Db 8526 TANTTTTGTATTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGT 8474

RESULT 84
AAZ50904/c
ID AAZ50904 standard; DNA; 17590 BP.
XX
XX AC AAZ50904;
XX
XX DT 31-MAY-2000 (first entry)
XX
XX DE Human TBC-1 partial genomic DNA comprising 5' end sequence.
XX
XX KW TBC-1; human; biallelic marker; chromosome 4; cell cycle regulator; SNP;
XX Single nucleotide polymorphism; tissue differentiation; prostate cancer;
XX linkage analysis; genetic map; detection; diagnosis; genotyping;
XX transgenic animal; screening; ds.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX misc_signal 1..2000
XX /tag= a
XX /note= "5', regulatory region"
XX /number= 1
XX /tag= b
XX /number= 1
XX /tag= c
XX /number= 1
XX /tag= d
XX /note= "Amplification of amplicon 99-430"
XX /tag= e
XX /bound_moiety= "Primer B1"
XX /note= "Amplification of amplicon 99-430"
XX /tag= f
XX /bound_moiety= "Primer D1"
XX /note= "Microsequencing of marker 99-430-352"
XX 9482..9506
XX /tag= g
XX /bound_moiety= "Probe P1"
XX /note= "Detection of Biallelic marker 99-430-352"
XX 9494
XX /tag= h
XX /note= "Ambiguity base 'R' corresponds to 'A' in allele-1
XX and 'G' in allele-2 of biallelic marker 99-430-352"
XX complement (9495..9513)
XX /tag= i
XX /bound_moiety= "Primer E1"
XX /note= "Microsequencing of marker 99-430-352"
XX complement (9828..9845)
XX /tag= j
XX /bound_moiety= "Primer C1"
XX /note= "Amplification of amplicon 99-430"
XX 12292..12373
XX /tag= k
XX /number= 1 bis
XX 12374..12739
XX /tag= l
XX /number= 1 bis
XX 12740..13249
XX /tag= m
XX /number= 2
XX 13250..17590
XX /tag= n
XX /number= 2
XX
XX WO200008209-A2.
XX
XX 17-FEB-2000.
XX
XX 06-AUG-1999; 99WO-IB01444.
XX
XX 07-AUG-1998; 98US-0095653.
XX
XX (GEST ) GENSET.
XX
XX PI Blumenfeld M, Bougueleret L, Chumakov I;
XX
XX WPI: 2000-205736/18.
XX
XX New isolated human TBC-1 nucleic acids, useful for developing products
XX for the diagnosis and treatment of disorders involving cell
XX proliferation, particularly prostate cancer
XX
XX Claim 1; Page 93-100; 166pp; English.
XX
XX The present sequence is the partial genomic DNA of human TBC-1 gene,
XX comprising the 5' regulatory region, exons 1, 1bis and 2. TBC-1 gene is
XX mapped to a candidate region of prostate cancer on chromosome 4. Single
XX nucleotide polymorphism (SNP) is located within the biallelic marker
XX region 99-430-352, localised in intron 1 of TBC-1 genomic DNA.
XX TBC-1 gene is involved in the regulation of cell cycle and tissue
XX differentiation in mammals. An alteration of TBC-1 sequence may be
XX associated with a pathological condition, resulting in abnormal cell
XX proliferation leading to cancer, e.g. prostate cancer. The biallelic
XX markers can be used for generation of genetic maps, linkage analysis and
XX association studies. TBC-1 sequence can be used for detection,
XX diagnosis, genotyping, production of transgenic animals and screening

```

CC of compounds for use in therapy.

XX Sequence 17590 BP; 4760 A; 3776 C; 4104 G; 4919 T; 31 other;

Query Match 0.2%; Score 53; DB 21; Length 17590;
Best Local Similarity 100.0%; Pred. No. 1.6e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagatagatgggtttccaccatgttggccagctggt 9067
|||||
Db 11661 TAATTTTGTATTTTCTAGATAGATGGGTTCACCATGTTGGCCAGCTGCT 11609
|||||

RESULT 85

AAF97862/C

ID AAF97862 standard; DNA; 22081 BP.

XX AAF97862;

XX 31-MAY-2001 (first entry)

XX Human neuroblastoma cell line NB-1 lp36 nucleotide sequence SEQ ID NO:76.

DE Human; chromosome 1; lp36; neuroblastoma cell line; NB-1; anticancer;
XX tumour suppressor; human lp36 homozygosity deletion domain; tumour;
KW diagnosis; ds.

XX Homo sapiens.

XX WO200116311-A1.

XX 08-MAR-2001.

XX 31-AUG-2000; 2000WO-JP05930.

XX 31-AUG-1999; 99JP-0245962.

XX 09-MAY-2000; 2000JP-0136266.

XX (HISM) HISAMITSU PHARM CO LTD.
XX (CHIB-) CHIBA PREFECTURE.

XX Nakagawara A;

XX WPI; 2001-226686/23.

XX Human lp36 homozygosity deletion domain from the 36-position of first
PT chromosome short arm in human neuroblastoma cell lines, applicable e.g.
PT in gene diagnosis of tumors as well as in developing anti-cancer drugs

XX Example 8; Page 149-158; 2266pp; Japanese.

XX The present invention describes a homozygosity deletion domain
CC co-existing in the 36-position of the first chromosome short arm (lp36)
CC in human neuroblastoma. Also described are base sequences from the lp36
CC position of human neuroblastoma cell lines (NB-1 and MASS-NB-SCH-1),
CC which are tumour suppressor genes in human neuroblastoma. The genes are
CC tumour suppressor genes, base sequence data of which are applicable as
CC tumour markers and reagents in studying mechanism of tumour body
CC formation, and gene diagnosis of tumours as well as in developing
CC anti-cancer drugs. AAF97787 to AAF97829 represent PCR primers used in
CC the exemplification of the present invention, and AAF97830 to AAF97874
CC represent sequences given in the exemplification of the present
CC invention.

XX Sequence 22081 BP; 5910 A; 5508 C; 5430 G; 5233 T; 0 other;

Query Match 0.2%; Score 53; DB 22; Length 22081;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagatagatgggtttccaccatgttggccagctggt 9067
|||||
Db 9199 TAATTTTGTATTTTCTAGATAGATGGGTTCACCATGTTGGCCAGCTGCT 9147
|||||

RESULT 86

AAT17455

ID AAT17455 standard; cDNA; 24025 BP.

XX AAT17455;

XX 07-OCT-1996 (first entry)

XX Mutated BRCA1 genomic sequence from sample set MSKCC family 19921.

XX Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.

XX Homo sapiens.

XX Key Location/Qualifiers
XX 256..355

FT /tag= a

FT /note= "exon 1"

FT 356..1512

FT /tag= b

FT /note= "intron 1"

FT 1295

FT /tag= c

FT /note= "known polymorphic site"

FT 1513..1611

FT /tag= d

FT /note= "exon 2"

FT 1612..2206

FT /tag= e

FT /note= "intron 2"

FT 1925..1937

FT /tag= f

FT /note= "indefinite interval within intron 2"

FT 2141

FT /tag= g

FT /note= "known polymorphic site"

FT 2207..2260

FT /tag= h

FT /note= "exon 3"

FT 2261..2677

FT /tag= i

FT /note= "intron 3"

FT 2569..2581

FT /tag= j

FT /note= "indefinite interval within intron 3"

FT 2678..2788

FT /tag= k

FT /note= "exon 4"

FT 2725

FT /tag= l

FT /note= "known polymorphic site"

FT 2789..3328

FT /tag= m

FT /note= "intron 4"

FT 3063..3075

FT /tag= n

FT /note= "indefinite interval within intron 4"

FT 3329..3406

FT /tag= o

FT /note= "exon 5"

FT 3407..3813

FT /tag= p

FT /note= "intron 5"

FT 3598..3610

FT /tag= q

FT /note= "indefinite interval within intron 5"

```
FT misc_feature 3653 /*tag= r
FT /note= "known polymorphic site"
FT 3814..3902 /*tag= s
FT /*tag= "exon 6"
FT intron 3903..4224 /*tag= t
FT /note= "intron 6"
FT mutation 4223 /*tag= u
FT /note= "Site of 1 nucleotide deletion"
FT 4076..4088 /*tag= v
FT /note= "indefinite interval within intron 6"
FT exon 4225..4364 /*tag= w
FT /*tag= "exon 7"
FT intron 4365..6571 /*tag= x
FT /note= "intron 7"
FT misc_feature 4391..4392 /*tag= y
FT /note= "known polymorphic site"
FT misc_feature 4602..4614 /*tag= z
FT /note= "indefinite interval within intron 7"
FT misc_feature 6538 /*tag= aa
FT /note= "known polymorphic site"
FT exon 6572..6677 /*tag= ab
FT /note= "exon 8"
FT intron 6678..9163 /*tag= ac
FT /note= "intron 8"
FT misc_feature 6823 /*tag= ad
FT /note= "known polymorphic site"
FT misc_feature 9106 /*tag= ae
FT /note= "known polymorphic site"
FT exon 9164..9209 /*tag= af
FT /note= "exon 9"
FT misc_feature 9207 /*tag= ag
FT /note= "known polymorphic site"
FT intron 9210..10530 /*tag= ah
FT /note= "intron 9"
FT misc_feature 9376 /*tag= ai
FT /note= "known polymorphic site"
FT exon 10531..10607 /*tag= aj
FT /note= "exon 10"
FT intron 10608..11597 /*tag= ak
FT /note= "intron 10"
FT misc_feature 11384..11396 /*tag= al
FT /note= "indefinite interval within intron 10"
FT exon 11598..15023 /*tag= am
FT /note= "exon 11"
FT misc_feature 11908 /*tag= an
FT /note= "known polymorphic site"
FT misc_feature 11994 /*tag= ao
FT /note= "known polymorphic site"
FT misc_feature 12952
```

```
FT /*tag= ap
FT /note= "known polymorphic site"
FT 13004 /*tag= aq
FT /*tag= "known polymorphic site"
FT misc_feature 13009 /*tag= ar
FT /note= "known polymorphic site"
FT 13048 /*tag= as
FT /*tag= "known polymorphic site"
FT 13238 /*tag= at
FT /note= "known polymorphic site"
FT 13448 /*tag= au
FT /note= "known polymorphic site"
FT 13539 /*tag= av
FT /note= "known polymorphic site"
FT 13951 /*tag= aw
FT /note= "known polymorphic site"
FT 14041 /*tag= ax
FT /note= "known polymorphic site"
FT 14046 /*tag= ay
FT /note= "known polymorphic site"
FT 14475 /*tag= az
FT /note= "known polymorphic site"
FT 14874 /*tag= ba
FT /note= "known polymorphic site"
FT 14891 /*tag= bb
FT /note= "known polymorphic site"
FT 14966 /*tag= bc
FT /note= "known polymorphic site"
FT intron 15024..15424 /*tag= bd
FT /note= "intron 11"
FT 15284 /*tag= be
FT /note= "known polymorphic site"
FT exon 15425..15511 /*tag= bf
FT /note= "exon 12"
FT intron 15512..15952 /*tag= bg
FT /note= "intron 12"
FT misc_feature 15647..15659 /*tag= bh
FT /note= "indefinite interval within intron 12"
FT exon 15953..16126 /*tag= bi
FT /note= "exon 13"
FT 16077 /*tag= bj
```

Query Match 0.2%; Score 53; DB 17; Length 24025;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtatttttagtagagatgggtttccaccatgttggccaggctggt 9067
|||||

Db 10222 taattttgtatttttagtagagatgggtttccaccatgttggccaggctggt 10274

RESULT 87
AAT17515


```
FT      /*tag= ar
FT      /note= "known polymorphic site"
FT      13237
FT      /*tag= as
FT      /note= "known polymorphic site"
FT      13447
FT      /*tag= at
FT      /note= "known polymorphic site"
FT      13538
FT      /*tag= au
FT      /note= "known polymorphic site"
FT      13950
FT      /*tag= av
FT      /note= "known polymorphic site"
FT      14040
FT      /*tag= aw
FT      /note= "known polymorphic site"
FT      14045
FT      /*tag= ax
FT      /note= "known polymorphic site"
FT      14474
FT      /*tag= ay
FT      /note= "known polymorphic site"
FT      14873
FT      /*tag= az
FT      /note= "known polymorphic site"
FT      14890
FT      /*tag= ba
FT      /note= "known polymorphic site"
FT      14965
FT      /*tag= bb
FT      /note= "known polymorphic site"
FT      15023..15423
FT      /*tag= bc
FT      /note= "intron 11"
FT      15283
FT      /*tag= bd
FT      /note= "known polymorphic site"
FT      15424..15510
FT      /*tag= be
FT      /note= "exon 12"
FT      15511..15951
FT      /*tag= bf
FT      /note= "intron 12"
FT      15646..15658
FT      /*tag= bg
FT      /note= "indefinite interval within intron 12"
FT      15952..16125
FT      /*tag= bh
FT      /note= "exon 13"
FT      16076
FT      /*tag= bi
FT      /note= "known polymorphic site"
FT      16126..16564
```

```
Query Match      0.2%; Score 53; DB 17; Length 24025;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 9015 taattttgtatttttagtagagatgggtttccaccatgttggccagctggt 9067
|||||
Db 10222 taattttgtatttttagtagagatgggtttccaccatgttggccagctggt 10274
```

RESULT 88

AAT32612
ID AAT32612 standard; DNA; 24026 BP.

XX AAT32612;

AC AAT32612;

XX 19-NOV-1996 (first entry)

XX BRCA1, human breast and ovarian cancer predisposing gene.

```
XX BRCA1: breast cancer; ovary cancer; predisposing gene; chromosome 17q;
KW susceptibility gene; diagnosis; prognosis; gene therapy; gene mapping;
KW marker; testis; thymus; exon; intron; ds.
XX Homo sapiens.
FH Key
FT Location/Qualifiers
FT 1..55
FT /*tag= a
FT 56..155
FT /*tag= b
FT 156..1512
FT /*tag= c
FT 1513..1611
FT /*tag= d
FT 1612..2206
FT /*tag= e
FT /note= "n at 1925-1937 represent an indefinite
FT interval within the intron"
FT 2207..2260
FT /*tag= f
FT 2261..2677
FT /*tag= g
FT /note= "n at 2569-2581 represent an indefinite
FT interval within the intron"
FT 2678..2788
FT /*tag= h
FT 2789..3328
FT /*tag= i
FT /note= "n at 3063-3075 represent an indefinite
FT interval within the intron"
FT 3329..3406
FT /*tag= j
FT 3407..3813
FT /*tag= k
FT /note= "n at 3598-3610 represent an indefinite
FT interval within the intron"
FT 3814..3902
FT /*tag= l
FT 3903..4224
FT /*tag= m
FT /note= "n at 4076-4088 represent an indefinite
FT interval within the intron"
FT 4225..4364
FT /*tag= n
FT 4365..6571
FT /*tag= o
FT /note= "n at 4602-4614 represent an indefinite
FT interval within the intron"
FT 6572..6677
FT /*tag= p
FT 6678..9163
FT /*tag= q
FT 9164..9207
FT /*tag= r
FT 9208..10530
FT /*tag= s
FT 10531..10607
FT /*tag= t
FT 10608..11597
FT /*tag= u
FT /note= "n at 11383-11396 represent an indefinite
FT interval within the intron"
FT 11598..15023
FT /*tag= v
FT 15024..15424
FT /*tag= w
FT 15425..15511
FT /*tag= x
FT 15512..15952
FT /*tag= y
FT /note= "n at 15647-15659 represent an indefinite
```



```
FT      /*tag= aw          /note= "known polymorphic site"
FT      14046
FT      /*tag= ax          /note= "known polymorphic site"
FT      14475
FT      /*tag= ay          /note= "known polymorphic site"
FT      14874
FT      /*tag= az          /note= "known polymorphic site"
FT      14891
FT      /*tag= ba          /note= "known polymorphic site"
FT      14966
FT      /*tag= bb          /note= "known polymorphic site"
FT      15024..15424
FT      /*tag= bc          /note= "known polymorphic site"
FT      15284
FT      /*tag= bd          /note= "C to A mutation at known polymorphic site"
FT      15425..15511
FT      /*tag= be          /note= "known polymorphic site"
FT      15512..15952
FT      /*tag= bf          /note= "known polymorphic site"
FT      15847..15859
FT      /*tag= bg          /note= "indefinite interval within intron 12"
FT      15953..16126
FT      /*tag= bh          /note= "known polymorphic site"
FT      16077
FT      /*tag= bi          /note= "known polymorphic site"
FT      16127..16565
FT      /*tag= bj          /note= "known polymorphic site"

Query Match      0.28; Score 53; DB 17; Length 24026;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  9015 taattttgtatttttagtagatggggtttccaccatgttgccagctggt 9067
      |||||
Db  10223 taattttgtatttttagtagatggggtttccaccatgttgccagctggt 10275

RESULT 90
AAT17513
ID  AAT17513 standard; CDNA; 24026 BP.
XX
AC  AAT17513;
XX
DT  04-OCT-1996 (first entry)
XX
DE  Mutated BRCA1 genomic sequence from PM05.
XX
KW  Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW  antibody production; germline alteration; probe; lesion neoplasia; human;
KW  gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
XX
OS  Homo sapiens.
XX
FH  Key      Location/Qualifiers
FT      exon      256..355
FT      /*tag= a
FT      /note= "exon 1"
FT      intron      356..1512
FT      /*tag= b
FT      /note= "intron 1"
```

```
FT      misc_feature      1295
FT      /*tag= c          /note= "known polymorphic site"
FT      1513..1611
FT      /*tag= d          /note= "exon 2"
FT      1612..2206
FT      /*tag= e          /note= "intron 2"
FT      1925..1937
FT      /*tag= f          /note= "indefinite interval within intron 2"
FT      2141
FT      /*tag= g          /note= "known polymorphic site"
FT      2207..2260
FT      /*tag= h          /note= "exon 3"
FT      2261..2677
FT      /*tag= i          /note= "intron 3"
FT      2589..2581
FT      /*tag= j          /note= "indefinite interval within intron 3"
FT      2678..2788
FT      /*tag= k          /note= "exon 4"
FT      2725
FT      /*tag= l          /note= "known polymorphic site"
FT      2789..3328
FT      /*tag= m          /note= "intron 4"
FT      3083..3075
FT      /*tag= n          /note= "indefinite interval within intron 4"
FT      3329..3406
FT      /*tag= o          /note= "exon 5"
FT      3407..3813
FT      /*tag= p          /note= "intron 5"
FT      3598..3610
FT      /*tag= q          /note= "indefinite interval within intron 5"
FT      3653
FT      /*tag= r          /note= "known polymorphic site"
FT      3814..3902
FT      /*tag= s          /note= "exon 6"
FT      3903..4224
FT      /*tag= t          /note= "intron 6"
FT      4076..4088
FT      /*tag= u          /note= "indefinite interval within intron 6"
FT      4225..4364
FT      /*tag= v          /note= "exon 7"
FT      4365..6571
FT      /*tag= w          /note= "intron 7"
FT      4391..4392
FT      /*tag= x          /note= "known polymorphic site"
FT      4602..4614
FT      /*tag= y          /note= "indefinite interval within intron 7"
FT      6538
FT      /*tag= z          /note= "known polymorphic site"
FT      6572..6677
```



```
FT      /*tag= aa
FT      /note= "exon 8"
FT      6678..9163
FT      /*tag= ab
FT      /note= "intron 8"
FT      6823
FT      /*tag= ac
FT      /note= "known polymorphic site"
FT      9106
FT      /*tag= ad
FT      /note= "known polymorphic site"
FT      9164..9209
FT      /*tag= ae
FT      /note= "exon 9"
FT      9207
FT      /*tag= af
FT      /note= "known polymorphic site"
FT      9210..10530
FT      /*tag= ag
FT      /note= "intron 9"
FT      9376
FT      /*tag= ah
FT      /note= "known polymorphic site"
FT      10531..10607
FT      /*tag= ai
FT      /note= "exon 10"
FT      10608..11597
FT      /*tag= aj
FT      /note= "intron 10"
FT      11384..11396
FT      /*tag= ak
FT      /note= "indefinite interval within intron 10"
FT      11598..15023
FT      /*tag= al
FT      /note= "exon 11"
FT      11908
FT      /*tag= am
FT      /note= "known polymorphic site"
FT      11994
FT      /*tag= an
FT      /note= "known polymorphic site"
FT      12952
FT      /*tag= ao
FT      /note= "known polymorphic site"
FT      13004
FT      /*tag= ap
FT      /note= "known polymorphic site"
FT      13009
FT      /*tag= aq
FT      /note= "known polymorphic site"
FT      13048
FT      /*tag= ar
FT      /note= "known polymorphic site"
FT      13238
FT      /*tag= as
FT      /note= "known polymorphic site"
FT      13448
FT      /*tag= at
FT      /note= "known polymorphic site"
FT      13539
FT      /*tag= au
FT      /note= "known polymorphic site"
FT      13951
FT      /*tag= av
FT      /note= "known polymorphic site"
FT      14041
FT      /*tag= aw
FT      /note= "known polymorphic site"
FT      14046
FT      /*tag= ax
FT      /note= "known polymorphic site"
FT      14475
FT      /*tag= ay
```

```
FT      misc_feature
FT      /note= "known polymorphic site"
FT      14874
FT      /*tag= az
FT      /note= "known polymorphic site"
FT      14891
FT      /*tag= ba
FT      /note= "known polymorphic site"
FT      14966
FT      /*tag= bb
FT      /note= "known polymorphic site"
FT      15024..15424
FT      /*tag= bc
FT      /note= "intron 11"
FT      15284
FT      /*tag= bd
FT      /note= "known polymorphic site"
FT      15425..15511
FT      /*tag= be
FT      /note= "exon 12"
FT      15512..15952
FT      /*tag= bf
FT      /note= "intron 12"
FT      15647..15659
FT      /*tag= bg
FT      /note= "indefinite interval within intron 12"
FT      15953..16126
FT      /*tag= bh
FT      /note= "exon 13"
FT      16077
FT      /*tag= bi
FT      /note= "known polymorphic site"
FT      16127..16565
FT      /*tag= bj
```

Query Match 0.2%; Score 53; DB 17; Length 24026;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 9015 taattttgtatttttagtagagatggggtttcaccatgttgccagctggt 9067
      |||||
Db 10223 taattttgtatttttagtagagatggggtttcaccatgttgccagctggt 10275
```

RESULT 91

AAT17514
ID AAT17514 standard; cDNA; 24026 BP.

XX AC AAT17514;

XX DT 04-OCT-1996 (first entry)

XX DE Mutated BRCA1 genomic sequence from PM11.

XX KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
XX KW antibody production; germline alteration; probe; lesion neoplasia; human;
XX KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT exon 256..355

FT /*tag= a

FT /note= "exon 1"

FT intron 356..1512

FT /*tag= b

FT /note= "intron 1"

FT misc_feature 1295

FT /*tag= c

FT /note= "known polymorphic site"

FT exon 1513..1611

FT /*tag= d

FT /note= "exon 2"

FT intron 1612..2206

```
FT FT      /*tag= e  
FT FT      /note= "intron 2"  
FT FT      1925..1937  
FT FT      /*tag= f  
FT FT      /note= "indefinite interval within intron 2"  
FT FT      2141  
FT FT      /*tag= g  
FT FT      /note= "known polymorphic site"  
FT FT      2207..2260  
FT FT      /*tag= h  
FT FT      /note= "exon 3"  
FT FT      2261..2677  
FT FT      /*tag= i  
FT FT      /note= "intron 3"  
FT FT      2569..2581  
FT FT      /*tag= j  
FT FT      /note= "indefinite interval within intron 3"  
FT FT      2678..2788  
FT FT      /*tag= k  
FT FT      /note= "exon 4"  
FT FT      2725  
FT FT      /*tag= l  
FT FT      /note= "known polymorphic site"  
FT FT      2789..3328  
FT FT      /*tag= m  
FT FT      /note= "intron 4"  
FT FT      3063..3075  
FT FT      /*tag= n  
FT FT      /note= "indefinite interval within intron 4"  
FT FT      3329..3406  
FT FT      /*tag= o  
FT FT      /note= "exon 5"  
FT FT      3407..3813  
FT FT      /*tag= p  
FT FT      /note= "intron 5"  
FT FT      3596..3610  
FT FT      /*tag= q  
FT FT      /note= "indefinite interval within intron 5"  
FT FT      3653  
FT FT      /*tag= r  
FT FT      /note= "known polymorphic site"  
FT FT      3814..3902  
FT FT      /*tag= s  
FT FT      /note= "exon 6"  
FT FT      3903..4224  
FT FT      /*tag= t  
FT FT      /note= "intron 6"  
FT FT      4076..4088  
FT FT      /*tag= u  
FT FT      /note= "indefinite interval within intron 6"  
FT FT      4225..4364  
FT FT      /*tag= v  
FT FT      /note= "exon 7"  
FT FT      4365..6571  
FT FT      /*tag= w  
FT FT      /note= "intron 7"  
FT FT      4391..4392  
FT FT      /*tag= x  
FT FT      /note= "known polymorphic site"  
FT FT      4602..4614  
FT FT      /*tag= y  
FT FT      /note= "indefinite interval within intron 7"  
FT FT      6538  
FT FT      /*tag= z  
FT FT      /note= "known polymorphic site"  
FT FT      6572..6677  
FT FT      /*tag= aa  
FT FT      /note= "exon 8"  
FT FT      6678..9163  
FT FT      /*tag= ab  
FT FT      /note= "intron 8"  
FT FT      6823  
FT FT      /*tag= ac  
FT FT      /note= "known polymorphic site"  
FT FT      9106  
FT FT      /*tag= ad  
FT FT      /note= "known polymorphic site"  
FT FT      9164..9209  
FT FT      /*tag= ae  
FT FT      /note= "exon 9"  
FT FT      9207  
FT FT      /*tag= af  
FT FT      /note= "known polymorphic site"  
FT FT      9210..10530  
FT FT      /*tag= ag  
FT FT      /note= "intron 9"  
FT FT      9376  
FT FT      /*tag= ah  
FT FT      /note= "known polymorphic site"  
FT FT      10531..10607  
FT FT      /*tag= ai  
FT FT      /note= "exon 10"  
FT FT      10608..11597  
FT FT      /*tag= aj  
FT FT      /note= "intron 10"  
FT FT      11384..11396  
FT FT      /*tag= ak  
FT FT      /note= "indefinite interval within intron 10"  
FT FT      11598..15023  
FT FT      /*tag= al  
FT FT      /note= "exon 11"  
FT FT      11908  
FT FT      /*tag= am  
FT FT      /note= "known polymorphic site"  
FT FT      11994  
FT FT      /*tag= an  
FT FT      /note= "known polymorphic site"  
FT FT      12952  
FT FT      /*tag= ao  
FT FT      /note= "known polymorphic site"  
FT FT      13004  
FT FT      /*tag= ap  
FT FT      /note= "known polymorphic site"  
FT FT      13009  
FT FT      /*tag= aq  
FT FT      /note= "known polymorphic site"  
FT FT      13048  
FT FT      /*tag= ar  
FT FT      /note= "known polymorphic site"  
FT FT      13238  
FT FT      /*tag= as  
FT FT      /note= "known polymorphic site"  
FT FT      13448  
FT FT      /*tag= at  
FT FT      /note= "known polymorphic site"  
FT FT      13539  
FT FT      /*tag= au  
FT FT      /note= "known polymorphic site"  
FT FT      13951  
FT FT      /*tag= av  
FT FT      /note= "known polymorphic site"  
FT FT      14041  
FT FT      /*tag= aw  
FT FT      /note= "known polymorphic site"  
FT FT      14046  
FT FT      /*tag= ax  
FT FT      /note= "known polymorphic site"  
FT FT      14475  
FT FT      /*tag= ay  
FT FT      /note= "known polymorphic site"  
FT FT      14874  
FT FT      /*tag= az  
FT FT      /note= "known polymorphic site"  
FT FT      14891  
FT FT      /*tag= ba  
FT FT      /note= "known polymorphic site"
```

```
FT misc_feature 14966 /*tag= bb /note= "known polymorphic site"
FT intron 15024..15424 /*tag= bc /note= "exon 3"
FT misc_feature 15284 /*tag= bd /note= "intron 11"
FT exon 15425..15511 /*tag= be /note= "intron 3"
FT intron 15512..15952 /*tag= bf /note= "indefinite interval within intron 3"
FT misc_feature 15647..15659 /*tag= bg /note= "exon 4"
FT exon 15953..16126 /*tag= bh /note= "exon 5"
FT intron 16077 /*tag= bi /note= "exon 6"
FT misc_feature 16127..16565 /*tag= bj /note= "exon 7"

Query Match 0.28; Score 53; DB 17; Length 24026;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagtagagatgggtttccaccattgtggccaggctggt 9067
|||||
Db 10223 taattttgtatttttagtagagatgggtttccaccattgtggccaggctggt 10275

RESULT 92
AAT17516
ID AAT17516 standard; cDNA; 24026 BP.
AC AAT17516;
XX
XX 04-OCT-1996 (first entry)
XX Mutated BRCA1 genomic sequence from PM16.
XX
XX Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
XX antibody production; germline alteration; probe; lesion neoplasia; human;
XX gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX exon 256..355
XX /*tag= a /note= "exon 1"
XX intron 356..1512
XX /*tag= b /note= "intron 1"
XX misc_feature 1295 /*tag= c /note= "known polymorphic site"
XX exon 1513..1611
XX /*tag= d /note= "exon 2"
XX intron 1612..2206
XX /*tag= e /note= "intron 2"
XX misc_feature 1925..1937 /*tag= f /note= "indefinite interval within intron 2"
XX 2141 /*tag= g
XX

FT exon /note= "known polymorphic site"
FT 2207..2260 /*tag= h
FT intron 2261..2677 /*tag= i
FT misc_feature 2569..2581 /*tag= j
FT exon 2678..2788 /*tag= k
FT misc_feature 2725 /*tag= l
FT intron 2789..3328 /*tag= m
FT misc_feature 3063..3075 /*tag= n
FT exon 3329..3406 /*tag= o
FT intron 3407..3813 /*tag= p
FT misc_feature 3598..3610 /*tag= q
FT misc_feature 3653 /*tag= r
FT exon 3814..3902 /*tag= s
FT intron 3903..4224 /*tag= t
FT misc_feature 4076..4088 /*tag= u
FT exon 4225..4364 /*tag= v
FT intron 4365..6571 /*tag= w
FT misc_feature 4391..4392 /*tag= x
FT misc_feature 4602..4614 /*tag= y
FT misc_feature 6538 /*tag= z
FT exon 6572..6677 /*tag= aa
FT intron 6678..9163 /*tag= ab
FT misc_feature 6823 /*tag= ac
FT misc_feature 9106 /*tag= ad
FT exon 9164..9209 /*tag= ae
FT /*tag= ae
FT /note= "exon 9"
```

```
FT misc_feature 9207 /*tag= af
FT /*note= "known polymorphic site"
FT 9210..10530
FT /*tag= ag
FT /*note= "intron 9"
FT 9376
FT /*tag= ah
FT /*note= "known polymorphic site"
FT 10531..10607
FT /*tag= ai
FT /*note= "exon 10"
FT 10608..11597
FT /*tag= aj
FT /*note= "intron 10"
FT 11384..11396
FT /*tag= ak
FT /*note= "indefinite interval within intron 10"
FT 11598..15023
FT /*tag= al
FT /*note= "exon 11"
FT 11908
FT /*tag= am
FT /*note= "known polymorphic site"
FT 11994
FT /*tag= an
FT /*note= "known polymorphic site"
FT 12952
FT /*tag= ao
FT /*note= "known polymorphic site"
FT 13004
FT /*tag= ap
FT /*note= "known polymorphic site"
FT 13009
FT /*tag= aq
FT /*note= "known polymorphic site"
FT 13048
FT /*tag= ar
FT /*note= "known polymorphic site"
FT 13238
FT /*tag= as
FT /*note= "known polymorphic site"
FT 13448
FT /*tag= at
FT /*note= "known polymorphic site"
FT 13539
FT /*tag= au
FT /*note= "known polymorphic site"
FT 13951
FT /*tag= av
FT /*note= "known polymorphic site"
FT 14041
FT /*tag= aw
FT /*note= "known polymorphic site"
FT 14046
FT /*tag= ax
FT /*note= "known polymorphic site"
FT 14475
FT /*tag= ay
FT /*note= "known polymorphic site"
FT 14874
FT /*tag= az
FT /*note= "known polymorphic site"
FT 14891
FT /*tag= ba
FT /*note= "known polymorphic site"
FT 14966
FT /*tag= bb
FT /*note= "known polymorphic site"
FT 15024..15424
FT /*tag= bc
FT /*note= "intron 11"
FT 15284
```

```
FT /*tag= bd
FT /*note= "known polymorphic site"
FT 15425..15511
FT /*tag= be
FT /*note= "exon 12"
FT 15512..15952
FT /*tag= bf
FT /*note= "intron 12"
FT 15647..15659
FT /*tag= bg
FT /*note= "indefinite interval within intron 12"
FT 15953..16126
FT /*tag= bh
FT /*note= "exon 13"
FT 16077
FT /*tag= bi
FT /*note= "known polymorphic site"
FT 16127..16365
FT /*tag= bj

Query Match 0.24; Score 53; DB 17; Length 24026;
Best Local Similarity 100.08; Pred. No. 1.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagtagagatgggtttccaccatgttgccaggctggt 9067
|||||
Db 10223 taattttgtatttttagtagagatgggtttccaccatgttgccaggctggt 10275

RESULT 93
AAT17517
ID AAT17517 standard; cDNA: 24026 BP.
XX
AC AAT17517;
XX
DT 04-OCT-1996 (first entry)
XX
DE Mutated BRCA1 genomic sequence from PMA02.1.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
XX
OS Homo sapiens.
XX
PH Key
FT exon Location/Qualifiers
FT /*tag= a
FT /*note= "exon 1"
FT 356..1512
FT /*tag= b
FT /*note= "intron 1"
FT 1295
FT mutation /*tag= c
FT /*note= "G to A mutation at known polymorphic site"
FT 1513..1611
FT /*tag= d
FT /*note= "exon 2"
FT 1612..2206
FT /*tag= e
FT /*note= "intron 2"
FT 1925..1937
FT /*tag= f
FT /*note= "indefinite interval within intron 2"
FT 2141
FT misc_feature /*tag= g
FT /*note= "known polymorphic site"
FT 2207..2260
FT /*tag= h
FT /*note= "exon 3"
FT 2261..2677
FT /*tag= i
FT /*note= "intron 3"
FT
```

FT	misc_feature	2569..2581	/*tag= j	/*tag= ah
FT		/*note= "indefinite interval within intron 3"	/*note= "known polymorphic site"	
FT	exon	2678..2788	/*tag= ai	10531..10607
FT		/*tag= k	/*tag= "exon 10"	
FT	misc_feature	2725	/*tag= aj	10608..11597
FT		/*tag= l	/*note= "intron 10"	
FT	intron	2789..3328	/*tag= ak	11384..11396
FT		/*note= "known polymorphic site"	/*note= "indefinite interval within intron 10"	
FT		/*tag= m		
FT	misc_feature	3063..3075	/*tag= ai	11598..15023
FT		/*tag= n	/*note= "exon 11"	
FT	exon	3329..3406	/*tag= am	11908
FT		/*tag= o	/*tag= "known polymorphic site"	
FT	intron	3407..3813	/*tag= an	11994
FT		/*tag= p	/*note= "known polymorphic site"	
FT	misc_feature	3598..3610	/*tag= ao	12952
FT		/*tag= q	/*note= "known polymorphic site"	
FT	misc_feature	3653	/*tag= ap	13004
FT		/*note= "indefinite interval within intron 5"	/*note= "known polymorphic site"	
FT	exon	3814..3902	/*tag= aq	13009
FT		/*tag= s	/*note= "known polymorphic site"	
FT	intron	3903..4224	/*tag= ar	13048
FT		/*tag= t	/*note= "known polymorphic site"	
FT	misc_feature	4076..4088	/*tag= as	13238
FT		/*tag= u	/*note= "known polymorphic site"	
FT	exon	4225..4364	/*tag= at	13448
FT		/*tag= v	/*note= "known polymorphic site"	
FT	intron	4365..6571	/*tag= au	13539
FT		/*tag= w	/*note= "known polymorphic site"	
FT	misc_feature	4391..4392	/*tag= av	13951
FT		/*tag= x	/*note= "known polymorphic site"	
FT	misc_feature	4602..4614	/*tag= aw	14041
FT		/*tag= y	/*note= "known polymorphic site"	
FT	misc_feature	6538	/*tag= ax	14046
FT		/*tag= z	/*note= "known polymorphic site"	
FT	exon	6572..6677	/*tag= ay	14475
FT		/*tag= aa	/*note= "known polymorphic site"	
FT	intron	6678..9163	/*tag= az	14874
FT		/*tag= ab	/*note= "known polymorphic site"	
FT	misc_feature	6823	/*tag= ba	14891
FT		/*tag= ac	/*note= "known polymorphic site"	
FT	misc_feature	9106	/*tag= bb	14966
FT		/*tag= ad	/*note= "known polymorphic site"	
FT	exon	9164..9209	/*tag= bc	15024..15424
FT		/*note= "exon 9"	/*note= "intron 11"	
FT	misc_feature	9207	/*tag= bd	15284
FT		/*tag= af	/*note= "known polymorphic site"	
FT	intron	9210..10530	/*tag= be	15425..15511
FT		/*tag= ag	/*note= "exon 12"	
FT	misc_feature	9376	/*tag= bf	15512..15952

```
FT misc_feature /note= "intron 12"
FT 15647..15659 /tag= bg
FT /note= "indefinite interval within intron 12"
FT 15953..16126 /tag= bh
FT /note= "exon 13"
FT 16077 /note= "exon 13"
FT /tag= bi
FT /note= "known polymorphic site"
FT 16127..16565
FT /tag= bj

Query Match 0.2%; Score 53; DB 17; Length 24026;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtatttttagatagatgggtttccaccatgttgccaggtggt 9067
|||||
Db 10223 taattttgtatttttagatagatgggtttccaccatgttgccaggtggt 10275

RESULT 94
AAT17518
ID AAT17518 standard; cDNA; 24026 BP.
XX
AC AAT17518;
XX
DT 04-OCR-1996 (first entry)
XX
DE Mutated BRCA1 genomic sequence from PMA03.1.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
XX
OS Homo sapiens.
XX
FH Key
FH exon 256..355 Location/Qualifiers
FT /tag= a
FT /note= "exon 1"
FT 356..1512 /tag= b
FT /note= "intron 1"
FT 1295 /tag= c
FT /tag= "known polymorphic site"
FT 1513..1611 /note= "known polymorphic site"
FT /tag= d
FT /note= "exon 2"
FT 1612..2206 /tag= e
FT /note= "intron 2"
FT 1925..1937 /tag= f
FT /note= "indefinite interval within intron 2"
FT 2141 /tag= g
FT /tag= "G to C mutation at known polymorphic site"
FT 2207..2260 /note= "known polymorphic site"
FT /tag= h
FT /note= "exon 3"
FT 2261..2677 /tag= i
FT /note= "intron 3"
FT 2569..2581 /tag= j
FT /note= "indefinite interval within intron 3"
FT 2678..2788 /tag= k
FT /note= "exon 4"
FT 2725 /tag= l

FT /tag= l
FT /note= "known polymorphic site"
FT 2789..3328 /tag= m
FT /note= "intron 4"
FT 3063..3075 /tag= n
FT /note= "indefinite interval within intron 4"
FT 3329..3406 /tag= o
FT /note= "exon 5"
FT 3407..3813 /tag= p
FT /note= "intron 5"
FT 3598..3610 /tag= q
FT /note= "indefinite interval within intron 5"
FT 3653 /tag= r
FT /note= "known polymorphic site"
FT 3814..3902 /tag= s
FT /note= "exon 6"
FT 3903..4224 /tag= t
FT /note= "intron 6"
FT 4076..4088 /tag= u
FT /note= "indefinite interval within intron 6"
FT 4225..4364 /tag= v
FT /note= "exon 7"
FT 4365..6571 /tag= w
FT /note= "intron 7"
FT 4391..4392 /tag= x
FT /note= "known polymorphic site"
FT 4602..4614 /tag= y
FT /note= "indefinite interval within intron 7"
FT 6538 /tag= z
FT /note= "known polymorphic site"
FT 6572..6677 /tag= aa
FT /note= "exon 8"
FT 6678..9163 /tag= ab
FT /note= "intron 8"
FT 6823 /tag= ac
FT /note= "known polymorphic site"
FT 9106 /tag= ad
FT /note= "known polymorphic site"
FT 9164..9209 /tag= ae
FT /note= "exon 9"
FT 9207 /tag= af
FT /note= "known polymorphic site"
FT 9210..10530 /tag= ag
FT /note= "intron 9"
FT 9376 /tag= ah
FT /note= "known polymorphic site"
FT 10531..10607 /tag= ai
FT /note= "exon 10"
FT 10608..11597 /tag= aj
FT
```


FT	exon	/note= "indefinite interval within intron 4"	FT	misc_feature	11908	/tag= am
FT		3329..3406	FT		/note= "known polymorphic site"	
FT		/tag= o	FT	misc_feature	11994	/tag= an
FT	intron	/note= "exon 5"	FT		/note= "known polymorphic site"	
FT		3407..3813	FT	misc_feature	12952	/tag= ao
FT		/tag= p	FT		/note= "known polymorphic site"	
FT	misc_feature	/note= "intron 5"	FT	misc_feature	13004	/tag= ap
FT		3598..3610	FT		/note= "known polymorphic site"	
FT		/tag= q	FT	misc_feature	13009	/tag= ag
FT	mutation	/note= "indefinite interval within intron 5"	FT		/note= "known polymorphic site"	
FT		3653	FT	misc_feature	13048	/tag= ar
FT		/tag= r	FT		/note= "known polymorphic site"	
FT	exon	/note= "A to G mutation at known polymorphic site"	FT	misc_feature	13238	/tag= as
FT		3814..3902	FT		/note= "known polymorphic site"	
FT		/tag= s	FT	misc_feature	13448	/tag= at
FT	intron	/note= "exon 6"	FT		/note= "known polymorphic site"	
FT		3903..4224	FT	misc_feature	13539	/tag= au
FT		/tag= t	FT		/note= "known polymorphic site"	
FT		/note= "intron 6"	FT	misc_feature	13951	/tag= av
FT	misc_feature	4076..4088	FT		/note= "known polymorphic site"	
FT		/tag= u	FT	misc_feature	14041	/tag= ax
FT	exon	/note= "indefinite interval within intron 6"	FT		/note= "known polymorphic site"	
FT		4225..4364	FT	misc_feature	14046	/tag= ay
FT		/tag= v	FT		/note= "known polymorphic site"	
FT	intron	4365..6571	FT	misc_feature	14475	/tag= az
FT		/tag= w	FT		/note= "known polymorphic site"	
FT		/note= "intron 7"	FT	misc_feature	14874	/tag= ba
FT	misc_feature	4391..4392	FT		/note= "known polymorphic site"	
FT		/tag= x	FT	misc_feature	14891	/tag= bb
FT		/note= "known polymorphic site"	FT		/note= "known polymorphic site"	
FT	misc_feature	4602..4614	FT	misc_feature	14966	/tag= bc
FT		/tag= y	FT		/note= "intron 11"	
FT		/note= "indefinite interval within intron 7"	FT	misc_feature	15284	/tag= bd
FT	misc_feature	6538	FT		/note= "known polymorphic site"	
FT		/tag= z	FT	exon	15425..15511	/tag= be
FT	exon	6572..6677	FT		/note= "exon 12"	
FT		/tag= aa	FT	intron	15512..15952	/tag= bf
FT	intron	6678..9163	FT		/note= "intron 12"	
FT		/tag= ab	FT	misc_feature	15647..15659	/tag= bg
FT		/note= "intron 8"	FT		/note= "indefinite interval within intron 12"	
FT	misc_feature	6823	FT	exon	15953..16126	/tag= bh
FT		/tag= ac	FT		/note= "exon 13"	
FT		/note= "known polymorphic site"	FT	misc_feature	16077	/tag= bi
FT	misc_feature	9106	FT		/note= "known polymorphic site"	
FT		/tag= ad	FT	intron	16127..16565	/tag= bj
FT	exon	9164..9209	FT			
FT		/tag= ae				
FT		/note= "exon 9"				
FT	misc_feature	9207				
FT		/tag= af				
FT	intron	9210..10530				
FT		/tag= ag				
FT		/note= "intron 9"				
FT	misc_feature	9376				
FT		/tag= ah				
FT	exon	10531..10607				
FT		/tag= ai				
FT		/note= "exon 10"				
FT	intron	10608..11597				
FT		/tag= aj				
FT		/note= "intron 10"				
FT	misc_feature	11384..11396				
FT		/tag= ak				
FT		/note= "indefinite interval within intron 10"				
FT	exon	11598..15023				
FT		/tag= al				
FT		/note= "exon 11"				

Query Match 0.2% Score 53; DB 17; Length 24026;

Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taatttttatttttagatagataggggtttccaccatgttgccagctgggt 9067
|||||
Db 10223 taatttttgatttttagatagataggggtttccaccatgttgccagctgggt 10275

RESULT 96

AAT17521

ID AAT17521 standard; cDNA; 24026 BP.

XX AAT17521;

XX 04-OCT-1996 (first entry)

DE Mutated BRCA1 genomic sequence from PMA08.1.

XX Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
XX Homo sapiens.

FH Key Location/Qualifiers

FT exon 256..355

FT /*tag= a

FT /note= "exon 1"

FT /*tag= b

FT /note= "intron 1"

FT /*tag= c

FT /note= "known polymorphic site"

FT /*tag= d

FT /note= "exon 2"

FT /*tag= e

FT /note= "intron 2"

FT /*tag= f

FT /note= "indefinite interval within intron 2"

FT /*tag= g

FT /note= "known polymorphic site"

FT /*tag= h

FT /note= "exon 3"

FT /*tag= i

FT /note= "intron 3"

FT /*tag= j

FT /note= "indefinite interval within intron 3"

FT /*tag= k

FT /note= "exon 4"

FT /*tag= l

FT /note= "known polymorphic site"

FT /*tag= m

FT /note= "intron 4"

FT /*tag= n

FT /note= "indefinite interval within intron 4"

FT /*tag= o

FT /note= "exon 5"

FT /*tag= p

FT /note= "intron 5"

FT misc_feature 3598..3610
FT /*tag= q
FT /note= "indefinite interval within intron 5"
FT misc_feature 3653
FT /*tag= r
FT /note= "known polymorphic site"
FT exon 3814..3902
FT /*tag= s
FT /note= "exon 6"
FT intron 3903..4224
FT /*tag= t
FT /note= "intron 6"
FT misc_feature 4076..4088
FT /*tag= u
FT /note= "indefinite interval within intron 6"
FT exon 4225..4364
FT /*tag= v
FT /note= "exon 7"
FT intron 4365..6571
FT /*tag= w
FT /note= "intron 7"
FT misc_feature 4391..4392
FT /*tag= x
FT /note= "known polymorphic site"
FT misc_feature 4602..4614
FT /*tag= y
FT /note= "indefinite interval within intron 7"
FT mutation 6538
FT /*tag= z
FT /note= "C to T mutation at known polymorphic site"
FT exon 6572..6677
FT /*tag= aa
FT /note= "exon 8"
FT intron 6678..9163
FT /*tag= ab
FT /note= "intron 8"
FT misc_feature 6823
FT /*tag= ac
FT /note= "known polymorphic site"
FT misc_feature 9106
FT /*tag= ad
FT /note= "known polymorphic site"
FT exon 9164..9209
FT /*tag= ae
FT /note= "exon 9"
FT misc_feature 9207
FT /*tag= af
FT /note= "known polymorphic site"
FT intron 9210..10530
FT /*tag= ag
FT /note= "intron 9"
FT misc_feature 9376
FT /*tag= ah
FT /note= "known polymorphic site"
FT exon 10531..10607
FT /*tag= ai
FT /note= "exon 10"
FT intron 10608..11597
FT /*tag= aj
FT /note= "intron 10"
FT misc_feature 11384..11396
FT /*tag= ak
FT /note= "indefinite interval within intron 10"
FT exon 11598..15023
FT /*tag= al
FT /note= "exon 11"
FT misc_feature 11908
FT /*tag= am
FT /note= "known polymorphic site"
FT misc_feature 11994
FT /*tag= an
FT /note= "known polymorphic site"
FT misc_feature 12952

```
FT      /*tag= ao
FT      /note= "known polymorphic site"
FT      misc_feature 13004
FT      /*tag= ap
FT      /note= "known polymorphic site"
FT      misc_feature 13009
FT      /*tag= aq
FT      /note= "known polymorphic site"
FT      misc_feature 13048
FT      /*tag= ar
FT      /note= "known polymorphic site"
FT      misc_feature 13238
FT      /*tag= as
FT      /note= "known polymorphic site"
FT      misc_feature 13448
FT      /*tag= at
FT      /note= "known polymorphic site"
FT      misc_feature 13539
FT      /*tag= au
FT      /note= "known polymorphic site"
FT      misc_feature 13951
FT      /*tag= av
FT      /note= "known polymorphic site"
FT      misc_feature 14041
FT      /*tag= aw
FT      /note= "known polymorphic site"
FT      misc_feature 14046
FT      /*tag= ax
FT      /note= "known polymorphic site"
FT      misc_feature 14475
FT      /*tag= ay
FT      /note= "known polymorphic site"
FT      misc_feature 14874
FT      /*tag= az
FT      /note= "known polymorphic site"
FT      misc_feature 14891
FT      /*tag= ba
FT      /note= "known polymorphic site"
FT      misc_feature 14966
FT      /*tag= bb
FT      /note= "known polymorphic site"
FT      intron 15024..15424
FT      /*tag= bc
FT      /note= "intron 11"
FT      misc_feature 15284
FT      /*tag= bd
FT      /note= "known polymorphic site"
FT      exon 15425..15511
FT      /*tag= be
FT      /note= "exon 12"
FT      intron 15512..15952
FT      /*tag= bf
FT      /note= "intron 12"
FT      misc_feature 15847..15859
FT      /*tag= bg
FT      /note= "indefinite interval within intron 12"
FT      exon 15953..16126
FT      /*tag= bh
FT      /note= "exon 13"
FT      misc_feature 16077
FT      /*tag= bi
FT      /note= "known polymorphic site"
FT      intron 16127..16565
FT      /*tag= bj

Query Match      0.28; Score 53; DB 17; Length 24026;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagtagagatgggtttccaccatgttggccagctggt 9067
      |||||||
Db 10223 taattttgtatttttagtagagatgggtttccaccatgttggccagctggt 10275
```

```
RESULT 97
AAT17522
ID AAT17522 standard; cDNA; 24026 BP.
XX
AC AAT17522;
XX
XX 04-OCT-1996 (first entry)
XX
XX Mutated BRCA1 genomic sequence from PMA08.2.
XX
XX Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT exon 256..355
FT      /*tag= a
FT      /note= "exon 1"
FT      intron 356..1512
FT      /*tag= b
FT      /note= "intron 1"
FT      misc_feature 1295
FT      /*tag= c
FT      /note= "known polymorphic site"
FT      exon 1513..1611
FT      /*tag= d
FT      /note= "exon 2"
FT      intron 1612..2206
FT      /*tag= e
FT      /note= "intron 2"
FT      misc_feature 1925..1937
FT      /*tag= f
FT      /note= "indefinite interval within intron 2"
FT      misc_feature 2141
FT      /*tag= g
FT      /note= "known polymorphic site"
FT      exon 2207..2260
FT      /*tag= h
FT      /note= "exon 3"
FT      intron 2261..2677
FT      /*tag= i
FT      /note= "intron 3"
FT      misc_feature 2569..2581
FT      /*tag= j
FT      /note= "indefinite interval within intron 3"
FT      exon 2678..2788
FT      /*tag= k
FT      /note= "exon 4"
FT      misc_feature 2725
FT      /*tag= l
FT      /note= "known polymorphic site"
FT      intron 2789..3328
FT      /*tag= m
FT      /note= "intron 4"
FT      misc_feature 3063..3075
FT      /*tag= n
FT      /note= "indefinite interval within intron 4"
FT      exon 3329..3406
FT      /*tag= o
FT      /note= "exon 5"
FT      intron 3407..3813
FT      /*tag= p
FT      /note= "intron 5"
FT      misc_feature 3598..3610
FT      /*tag= q
FT      /note= "indefinite interval within intron 5"
FT      misc_feature 3653
FT      /*tag= r
FT      /note= "known polymorphic site"
FT      exon 3814..3902
```

```
FT      /*tag= s
FT      /note= "exon 6"
FT      3903..4224
FT      /*tag= t
FT      /note= "intron 6"
FT      4076..44088
FT      /*tag= u
FT      /note= "indefinite interval within intron 6"
FT      4225..4364
FT      /*tag= v
FT      /note= "exon 7"
FT      4365..6571
FT      /*tag= w
FT      /note= "intron 7"
FT      4391..4392
FT      /*tag= x
FT      /note= "known polymorphic site"
FT      4602..4614
FT      /*tag= y
FT      /note= "indefinite interval within intron 7"
FT      6538
FT      /*tag= z
FT      /note= "known polymorphic site"
FT      6572..6677
FT      /*tag= aa
FT      /note= "exon 8"
FT      6678..9163
FT      /*tag= ab
FT      /note= "intron 8"
FT      6823
FT      /*tag= ac
FT      /note= "A to T mutation at known polymorphic site"
FT      9106
FT      /*tag= ad
FT      /note= "known polymorphic site"
FT      9164..9209
FT      /*tag= ae
FT      /note= "exon 9"
FT      9207
FT      /*tag= af
FT      /note= "known polymorphic site"
FT      9210..10530
FT      /*tag= ag
FT      /note= "intron 9"
FT      9376
FT      /*tag= ah
FT      /note= "known polymorphic site"
FT      10531..10607
FT      /*tag= ai
FT      /note= "exon 10"
FT      10608..11597
FT      /*tag= aj
FT      /note= "intron 10"
FT      11384..11396
FT      /*tag= ak
FT      /note= "indefinite interval within intron 10"
FT      11598..15023
FT      /*tag= al
FT      /note= "exon 11"
FT      11908
FT      /*tag= am
FT      /note= "known polymorphic site"
FT      11994
FT      /*tag= an
FT      /note= "known polymorphic site"
FT      12952
FT      /note= "known polymorphic site"
FT      13004
FT      /*tag= ao
FT      /note= "known polymorphic site"
FT      13004
FT      /*tag= ap
FT      /note= "known polymorphic site"
FT      13009
FT      /*tag= aq
```

```
FT      misc_feature
FT      /note= "known polymorphic site"
FT      13048
FT      /*tag= ar
FT      /note= "known polymorphic site"
FT      13238
FT      /*tag= as
FT      /note= "known polymorphic site"
FT      13448
FT      /*tag= at
FT      /note= "known polymorphic site"
FT      13539
FT      /*tag= au
FT      /note= "known polymorphic site"
FT      13951
FT      /*tag= av
FT      /note= "known polymorphic site"
FT      14041
FT      /*tag= aw
FT      /note= "known polymorphic site"
FT      14046
FT      /*tag= ax
FT      /note= "known polymorphic site"
FT      14475
FT      /*tag= ay
FT      /note= "known polymorphic site"
FT      14874
FT      /*tag= az
FT      /note= "known polymorphic site"
FT      14891
FT      /*tag= ba
FT      /note= "known polymorphic site"
FT      14966
FT      /*tag= bb
FT      /note= "known polymorphic site"
FT      15024..15424
FT      /*tag= bc
FT      /note= "intron 11"
FT      15284
FT      /*tag= bd
FT      /note= "known polymorphic site"
FT      15425..15511
FT      /*tag= be
FT      /note= "exon 12"
FT      15512..15952
FT      /*tag= bf
FT      /note= "intron 12"
FT      15647..15659
FT      /*tag= bg
FT      /note= "indefinite interval within intron 12"
FT      15953..16126
FT      /*tag= bh
FT      /note= "exon 13"
FT      16077
FT      /*tag= bi
FT      /note= "known polymorphic site"
FT      16127..16565
FT      /*tag= bj
```

Query Match 0.2%; Score 53; DB 17; Length 24026;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 9015 taattttgtatttttagtagagatgggtttaccattgtggccagctggt 9067
      |||||
Db 10223 taattttgtatttttagtagagatgggtttaccattgtggccagctggt 10275
```

```
RESULT 98
AAT17523
ID AAT17523 standard; cDNA; 24026 BP.
XX
AC AAT17523;
XX
```

```
DT 04-OCT-1996 (first entry)
XX Mutated BRCA1 genomic sequence from PMA09.2.
XX Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
PH exon 256..335
FT /*tag= a
FT /note= "exon 1"
FT intron 356..1512
FT /*tag= b
FT /note= "intron 1"
FT misc_feature 1295
FT /*tag= c
FT /note= "known polymorphic site"
FT exon 1513..1611
FT /*tag= d
FT intron 1612..2206
FT /note= "exon 2"
FT misc_feature 1925..1937
FT /note= "intron 2"
FT /tag= f
FT /note= "indefinite interval within intron 2"
FT misc_feature 2141
FT /*tag= g
FT /note= "known polymorphic site"
FT exon 2207..2260
FT /*tag= h
FT intron 2261..2677
FT /*tag= i
FT /note= "intron 3"
FT misc_feature 2569..2581
FT /*tag= j
FT /note= "indefinite interval within intron 3"
FT exon 2678..2788
FT /*tag= k
FT /note= "exon 4"
FT misc_feature 2725
FT /*tag= l
FT /note= "known polymorphic site"
FT intron 2789..3328
FT /*tag= m
FT misc_feature 3063..3075
FT /*tag= n
FT /note= "indefinite interval within intron 4"
FT exon 3329..3406
FT /*tag= o
FT /note= "exon 5"
FT intron 3407..3813
FT /*tag= p
FT /note= "intron 5"
FT misc_feature 3598..3610
FT /*tag= q
FT /note= "indefinite interval within intron 5"
FT misc_feature 3653
FT /*tag= r
FT /note= "known polymorphic site"
FT exon 3814..3902
FT /*tag= s
FT /note= "exon 6"
FT intron 3903..4224
FT /*tag= t
FT /note= "intron 6"
FT misc_feature 4076..4088
FT /*tag= u
FT /note= "indefinite interval within intron 6"
FT exon 4225..4364
FT /*tag= v
FT /note= "exon 7"
FT intron 4365..6571
FT /*tag= w
FT /note= "intron 7"
FT misc_feature 4391..4392
FT /*tag= x
FT /note= "known polymorphic site"
FT misc_feature 4602..4614
FT /*tag= y
FT /note= "indefinite interval within intron 7"
FT misc_feature 6538
FT /*tag= z
FT /note= "known polymorphic site"
FT exon 6572..6677
FT /*tag= aa
FT /note= "exon 8"
FT intron 6678..9163
FT /*tag= ab
FT /note= "intron 8"
FT misc_feature 6823
FT /*tag= ac
FT /note= "known polymorphic site"
FT misc_feature 9106
FT /*tag= ad
FT /note= "known polymorphic site"
FT exon 9164..9209
FT /*tag= ae
FT /note= "exon 9"
FT misc_feature 9207
FT /*tag= af
FT /note= "known polymorphic site"
FT intron 9210..10530
FT /*tag= ag
FT /note= "intron 9"
FT misc_feature 9376
FT /*tag= ah
FT /note= "known polymorphic site"
FT mutation 9376
FT /*tag= ah
FT /note= "known polymorphic site"
FT exon 10531..10607
FT /note= "T to C mutation at known polymorphic site"
FT intron 10608..11597
FT /*tag= ai
FT /note= "exon 10"
FT misc_feature 11384..11396
FT /*tag= aj
FT /note= "intron 10"
FT exon 11384..11396
FT /*tag= ak
FT /note= "indefinite interval within intron 10"
FT misc_feature 11598..15023
FT /*tag= al
FT /note= "exon 11"
FT misc_feature 11908
FT /*tag= am
FT /note= "known polymorphic site"
FT intron 11994
FT /*tag= an
FT /note= "known polymorphic site"
FT misc_feature 12952
FT /*tag= ao
FT /note= "known polymorphic site"
FT exon 13004
FT /*tag= ap
FT /note= "known polymorphic site"
FT intron 13009
FT /*tag= aq
FT /note= "known polymorphic site"
FT misc_feature 13048
FT /*tag= ar
FT /note= "known polymorphic site"
```

```

FT misc_feature 13238      as
FT /*tag= "known polymorphic site"
FT /note= "known polymorphic site"
FT misc_feature 13448      at
FT /*tag= "known polymorphic site"
FT /note= "known polymorphic site"
FT misc_feature 13539      au
FT /*tag= "known polymorphic site"
FT /note= "known polymorphic site"
FT misc_feature 13951      av
FT /*tag= "known polymorphic site"
FT /note= "known polymorphic site"
FT misc_feature 14041      aw
FT /*tag= "known polymorphic site"
FT /note= "known polymorphic site"
FT misc_feature 14046      ax
FT /*tag= "known polymorphic site"
FT /note= "known polymorphic site"
FT misc_feature 14475      ay
FT /*tag= "known polymorphic site"
FT /note= "known polymorphic site"
FT misc_feature 14874      az
FT /*tag= "known polymorphic site"
FT /note= "known polymorphic site"
FT misc_feature 14891      ba
FT /*tag= "known polymorphic site"
FT /note= "known polymorphic site"
FT misc_feature 14966      bb
FT /*tag= "known polymorphic site"
FT /note= "known polymorphic site"
FT intron 15024..15424      bc
FT /*tag= "intron 11"
FT /note= "intron 11"
FT misc_feature 15284      bd
FT /*tag= "known polymorphic site"
FT /note= "known polymorphic site"
FT exon 15425..15511      be
FT /*tag= "exon 12"
FT /note= "exon 12"
FT intron 15512..15952      bf
FT /*tag= "intron 12"
FT /note= "intron 12"
FT misc_feature 15647..15659      bg
FT /*tag= "indefinite interval within intron 12"
FT /note= "indefinite interval within intron 12"
FT exon 15953..16126      bh
FT /*tag= "exon 13"
FT /note= "exon 13"
FT misc_feature 16077      bi
FT /*tag=

Query Match 0.2%; Score 53; DB 17; Length 24026;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtattttagtagagatgggtttccaccatgttgcaggctggt 9067
|||||
Db 10223 taattttgtattttagtagagatgggtttccaccatgttgcaggctggt 10275
|||||

RESULT 99
AAT17524
ID AAT17524 standard; cDNA; 24026 BP.
XX AC AAT17524;
XX
DT 04-OCT-1996 (first entry)
XX
DE Mutated BRCA1 genomic sequence from PMAL3.1.
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
XX
OS Homo sapiens.
XX
FH Key
FT exon
FT /tag= a
FT /note= "exon 1"
FT intron
FT /tag= b
FT /note= "intron 1"
FT misc_feature
FT /tag= c
FT /note= "known polymorphic site"
FT exon
FT /tag= d
FT /note= "exon 2"
FT intron
FT /tag= e
FT /note= "intron 2"
FT misc_feature
FT /tag= f
FT /note= "indefinite interval within intron 2"
FT misc_feature
FT /tag= g
FT /note= "known polymorphic site"
FT exon
FT /tag= h
FT /note= "exon 3"
FT intron
FT /tag= i
FT /note= "intron 3"
FT misc_feature
FT /tag= j
FT /note= "indefinite interval within intron 3"
FT exon
FT /tag= k
FT /note= "exon 4"
FT misc_feature
FT /tag= l
FT /note= "known polymorphic site"
FT intron
FT /tag= m
FT /note= "intron 4"
FT misc_feature
FT /tag= n
FT /note= "indefinite interval within intron 4"
FT exon
FT /tag= o
FT /note= "exon 5"
FT intron
FT /tag= p
FT /note= "intron 5"
FT misc_feature
FT /tag= q
FT /note= "indefinite interval within intron 5"
FT misc_feature
FT /tag= r
FT /note= "known polymorphic site"
FT exon
FT /tag= s
FT /note= "exon 6"
FT intron
FT /tag= t
FT /note= "intron 6"
FT misc_feature
FT /tag= u
FT /note= "indefinite interval within intron 6"
FT exon
FT /tag= v
FT /note= "exon 7"
FT intron
FT /tag= w
FT /note= "intron 7"
```

```
FT misc_feature 4391..4392 /*tag= x /*tag= "known polymorphic site"
FT /*tag= x /*tag= "known polymorphic site"
FT misc_feature 4602..4614 /*tag= y /*tag= "known polymorphic site"
FT /*tag= y /*tag= "known polymorphic site"
FT misc_feature /*note= "indefinite interval within intron 7"
FT /*tag= z /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT exon 6572..6677 /*tag= aa /*tag= "known polymorphic site"
FT /*tag= aa /*tag= "known polymorphic site"
FT intron 6678..9163 /*tag= ab /*tag= "intron 8"
FT /*tag= ab /*tag= "intron 8"
FT misc_feature 6823 /*tag= ac /*tag= "known polymorphic site"
FT /*tag= ac /*tag= "known polymorphic site"
FT misc_feature 9106 /*tag= ad /*tag= "known polymorphic site"
FT /*tag= ad /*tag= "known polymorphic site"
FT exon 9164..9209 /*tag= ae /*tag= "exon 9"
FT /*tag= ae /*tag= "exon 9"
FT misc_feature 9207 /*tag= af /*tag= "known polymorphic site"
FT /*tag= af /*tag= "known polymorphic site"
FT intron 9210..10530 /*tag= ag /*tag= "intron 9"
FT /*tag= ag /*tag= "intron 9"
FT misc_feature 9376 /*tag= ah /*tag= "known polymorphic site"
FT /*tag= ah /*tag= "known polymorphic site"
FT exon 10531..10607 /*tag= ai /*tag= "exon 10"
FT /*tag= ai /*tag= "exon 10"
FT intron 10608..11597 /*tag= aj /*tag= "intron 10"
FT /*tag= aj /*tag= "intron 10"
FT misc_feature 11384..11396 /*tag= ak /*tag= "indefinite interval within intron 10"
FT /*tag= ak /*tag= "indefinite interval within intron 10"
FT exon 11598..15023 /*tag= al /*tag= "exon 11"
FT /*tag= al /*tag= "exon 11"
FT misc_feature 11908 /*tag= am /*tag= "known polymorphic site"
FT /*tag= am /*tag= "known polymorphic site"
FT misc_feature 11994 /*tag= an /*tag= "known polymorphic site"
FT /*tag= an /*tag= "known polymorphic site"
FT misc_feature 12952 /*tag= ao /*tag= "known polymorphic site"
FT /*tag= ao /*tag= "known polymorphic site"
FT misc_feature 13004 /*tag= ap /*tag= "known polymorphic site"
FT /*tag= ap /*tag= "known polymorphic site"
FT misc_feature 13009 /*tag= aq /*tag= "known polymorphic site"
FT /*tag= aq /*tag= "known polymorphic site"
FT misc_feature 13048 /*tag= ar /*tag= "known polymorphic site"
FT /*tag= ar /*tag= "known polymorphic site"
FT misc_feature 13238 /*tag= as /*tag= "known polymorphic site"
FT /*tag= as /*tag= "known polymorphic site"
FT misc_feature 13448 /*tag= at /*tag= "known polymorphic site"
FT /*tag= at /*tag= "known polymorphic site"
FT misc_feature 13539 /*tag= au /*tag= "known polymorphic site"
FT /*tag= au /*tag= "known polymorphic site"
FT misc_feature 13951
```

```
FT /*tag= av /*tag= "known polymorphic site"
FT /*note= "known polymorphic site"
FT misc_feature 14041 /*tag= aw /*tag= "known polymorphic site"
FT /*tag= aw /*tag= "known polymorphic site"
FT misc_feature 14046 /*tag= ax /*tag= "known polymorphic site"
FT /*tag= ax /*tag= "known polymorphic site"
FT misc_feature 14475 /*tag= ay /*tag= "known polymorphic site"
FT /*tag= ay /*tag= "known polymorphic site"
FT misc_feature 14874 /*tag= az /*tag= "known polymorphic site"
FT /*tag= az /*tag= "known polymorphic site"
FT misc_feature 14891 /*tag= ba /*tag= "known polymorphic site"
FT /*tag= ba /*tag= "known polymorphic site"
FT misc_feature 14966 /*tag= bb /*tag= "known polymorphic site"
FT /*tag= bb /*tag= "known polymorphic site"
FT intron 15024..15424 /*tag= bc /*tag= "intron 11"
FT /*tag= bc /*tag= "intron 11"
FT misc_feature 15284 /*tag= bd /*tag= "known polymorphic site"
FT /*tag= bd /*tag= "known polymorphic site"
FT exon 15425..15511 /*tag= be /*tag= "exon 12"
FT /*tag= be /*tag= "exon 12"
FT intron 15512..15952 /*tag= bf /*tag= "intron 12"
FT /*tag= bf /*tag= "intron 12"
FT misc_feature 15647..15659 /*tag= bg /*tag= "indefinite interval within intron 12"
FT /*tag= bg /*tag= "indefinite interval within intron 12"
FT exon 15953..16126 /*tag= bh /*tag= "exon 13"
FT /*tag= bh /*tag= "exon 13"
FT misc_feature 16077 /*tag= bi /*tag= "known polymorphic site"
FT /*tag= bi /*tag= "known polymorphic site"
FT intron 16127..16565 /*tag= bj /*tag= "exon 1"
FT /*tag= bj /*tag= "exon 1"

Query Match 0.2%; Score 53; DB 17; Length 24026;
Best Local Similarity 100.0%; Pred. No. 1.5e-08; Mismatches 0; Indels 0; Gaps 0;
Matches 53; Conservative 0;

Qy 9015 taattttgtatttttagtagagatgggtttccaccatgttgccagctggt 9067
|||||
Db 10223 taattttgtatttttagtagagatgggtttccaccatgttgccagctggt 10275

RESULT 100
AAT17526
ID AAT17526 standard; cDNA; 24026 BP.
XX XX
AC AAT17526;
XX XX
DT 04-OCT-1996 (first entry)
XX XX
DE Mutated BRCA1 genomic sequence from PMA15.2.
XX XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
XX XX
OS Homo sapiens.
XX XX
FH Key Location/Qualifiers
FT exon 256..355
FT /*tag= a
FT /*note= "exon 1"
```

```
FT intron 356..1512 /*tag= z
FT /*tag= b /note= "known polymorphic site"
FT /*note= "intron 1"
FT 1295 /*tag= c
FT /*tag= "known polymorphic site"
FT 1513..1611 /*tag= d
FT /*tag= "exon 2"
FT intron 1612..2206 /*tag= e
FT /*tag= "intron 2"
FT misc_feature 1925..1937 /*tag= f
FT /*note= "indefinite interval within intron 2"
FT misc_feature 2141 /*tag= g
FT /*note= "known polymorphic site"
FT exon 2207..2260 /*tag= h
FT /*tag= "exon 3"
FT intron 2261..2677 /*tag= i
FT /*note= "intron 3"
FT misc_feature 2569..2581 /*tag= j
FT /*note= "indefinite interval within intron 3"
FT exon 2678..2788 /*tag= k
FT /*tag= "exon 4"
FT misc_feature 2725 /*tag= l
FT /*note= "known polymorphic site"
FT intron 2789..3328 /*tag= m
FT /*tag= "intron 4"
FT misc_feature 3063..3075 /*tag= n
FT /*note= "indefinite interval within intron 4"
FT exon 3329..3406 /*tag= o
FT /*tag= "exon 5"
FT intron 3407..3813 /*tag= p
FT /*note= "intron 5"
FT misc_feature 3598..3610 /*tag= q
FT /*note= "indefinite interval within intron 5"
FT misc_feature 3653 /*tag= r
FT /*note= "known polymorphic site"
FT exon 3814..3902 /*tag= s
FT /*tag= "exon 6"
FT intron 3903..4224 /*tag= t
FT /*note= "intron 6"
FT misc_feature 4076..4088 /*tag= u
FT /*tag= "indefinite interval within intron 6"
FT exon 4225..4364 /*tag= v
FT /*tag= "exon 7"
FT intron 4365..6571 /*tag= w
FT /*tag= "intron 7"
FT misc_feature 4391..4392 /*tag= x
FT /*note= "known polymorphic site"
FT misc_feature 4602..4614 /*tag= y
FT /*note= "indefinite interval within intron 7"
FT misc_feature 6538

FT exon /*tag= z
FT /*note= "known polymorphic site"
FT 6572..6677 /*tag= aa
FT /*tag= "exon 8"
FT intron 6678..9163 /*tag= ab
FT /*note= "intron 8"
FT misc_feature 6823 /*tag= ac
FT /*note= "known polymorphic site"
FT misc_feature 9106 /*tag= ad
FT /*tag= "known polymorphic site"
FT exon 9164..9209 /*tag= ae
FT /*note= "exon 9"
FT misc_feature 9207 /*tag= af
FT /*note= "known polymorphic site"
FT intron 9210..10530 /*tag= ag
FT /*note= "intron 9"
FT misc_feature 9376 /*tag= ah
FT /*note= "known polymorphic site"
FT exon 10531..10607 /*tag= ai
FT /*note= "exon 10"
FT intron 10608..11597 /*tag= aj
FT /*note= "intron 10"
FT misc_feature 11384..11396 /*tag= ak
FT /*note= "indefinite interval within intron 10"
FT exon 11598..15023 /*tag= al
FT /*note= "exon 11"
FT misc_feature 11908 /*tag= am
FT /*note= "known polymorphic site"
FT misc_feature 11994 /*tag= an
FT /*note= "known polymorphic site"
FT misc_feature 12952 /*tag= ao
FT /*note= "known polymorphic site"
FT misc_feature 13004 /*tag= ap
FT /*note= "known polymorphic site"
FT misc_feature 13009 /*tag= aq
FT /*note= "known polymorphic site"
FT misc_feature 13048 /*tag= ar
FT /*note= "known polymorphic site"
FT misc_feature 13238 /*tag= as
FT /*note= "known polymorphic site"
FT misc_feature 13448 /*tag= at
FT /*note= "known polymorphic site"
FT misc_feature 13539 /*tag= au
FT /*note= "known polymorphic site"
FT misc_feature 13951 /*tag= av
FT /*note= "known polymorphic site"
FT misc_feature 14041 /*tag= aw
FT /*note= "known polymorphic site"
FT misc_feature 14046 /*tag= ax
```

```
FT misc_feature /note= "known polymorphic site"
FT 14475 /tag= ay
FT /note= "known polymorphic site"
FT 14874 /tag= az
FT /note= "known polymorphic site"
FT 14891 /tag= ba
FT /note= "known polymorphic site"
FT 14966 /tag= bb
FT /note= "known polymorphic site"
FT 15024..15424 /tag= bc
FT /note= "intron 11"
FT 15284 /tag= bd
FT /note= "known polymorphic site"
FT 15425..15511 /tag= be
FT /note= "exon 12"
FT 15512..15952 /tag= bf
FT /note= "intron 12"
FT 15647..15659 /tag= bg
FT /note= "indefinite interval within intron 12"
FT 15953..16126 /tag= bh
FT /note= "exon 13"
FT 16077 /tag= bi
FT /note= "known polymorphic site"
FT 16127..16365 /tag= bj
```

```
Query Match 0.2%; Score 53; DB 17; Length 24026;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 9015 taattttttagatagagatgggtttcaccatgttgccaggtggt 9067
|||||
Db 10223 taattttttagatagagatgggtttcaccatgttgccaggtggt 10275
```

Search completed: November 2, 2001, 20:56:04
Job time: 73784 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2001, 21:14:50 ; Search time 460.05 Seconds
(without alignments)
10972.268 Million cell updates/sec

Title: US-09-434-382-28
Perfect score: 26664
Sequence: 1 latcaggtgactgaattcta.....ttcgcaagttttttgaca 26664

Scoring table: OLIGO.NUC

Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 9465562 residues

Word size : 8

Total number of hits satisfying chosen parameters: 608797

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCRUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	0.2	36741	US-09-301-665-3	Sequence 3, Appl
2	60	0.2	9721	US-09-345-217-2	Sequence 2, Appl
3	60	0.2	152331	US-09-128-155-16	Sequence 16, Appl
4	60	0.2	176373	US-09-128-155-17	Sequence 17, Appl
5	59	0.2	31571	US-08-323-443B-1	Sequence 1, Appl
6	59	0.2	53526	US-08-658-136-2	Sequence 2, Appl
7	59	0.2	53577	US-08-658-136-1	Sequence 1, Appl
8	58	0.2	2688	US-08-909-965C-1	Sequence 1, Appl
9	57	0.2	1559	US-08-417-174-1	Sequence 1, Appl
10	57	0.2	1559	US-08-231-565A-1	Sequence 1, Appl
11	57	0.2	1559	US-09-007-961-1	Sequence 1, Appl
12	55	0.2	20303	US-08-370-975B-6	Sequence 6, Appl
13	55	0.2	26764	US-08-370-975B-1	Sequence 1, Appl
14	53	0.2	6769	US-08-480-784-20	Sequence 20, Appl
15	53	0.2	6769	US-08-483-553-20	Sequence 20, Appl
16	53	0.2	6769	US-08-487-002-20	Sequence 20, Appl
17	53	0.2	6769	US-08-483-554B-20	Sequence 20, Appl
18	53	0.2	6769	US-08-488-011B-20	Sequence 20, Appl
19	53	0.2	6769	US-08-850-727-20	Sequence 20, Appl
20	53	0.2	6769	PCT-US95-10202-20	Sequence 20, Appl
21	53	0.2	6769	PCT-US95-10203-20	Sequence 20, Appl
22	53	0.2	6769	PCT-US95-10220-20	Sequence 20, Appl
23	53	0.2	7720	US-09-318-448-5	Sequence 5, Appl
24	53	0.2	13865	US-09-009-217-11	Sequence 11, Appl
25	53	0.2	13865	US-09-009-656-11	Sequence 11, Appl
26	53	0.2	72928	US-09-009-913-1	Sequence 1, Appl
27	53	0.2	87350	US-08-781-891-79	Sequence 79, Appl

52	0.2	807	2	US-08-531-927B-9	Sequence 9, Appl
52	0.2	1776	2	US-08-531-927B-1	Sequence 1, Appl
52	0.2	1776	4	US-09-041-886-12	Sequence 12, Appl
52	0.2	17327	1	US-07-906-871-15	Sequence 15, Appl
51	0.2	1808	1	US-08-351-149-4	Sequence 4, Appl
51	0.2	1808	1	US-08-384-828-4	Sequence 4, Appl
51	0.2	1808	3	US-08-895-474-4	Sequence 4, Appl
51	0.2	2559	2	US-08-886-152-4	Sequence 4, Appl
51	0.2	8342	3	US-08-545-860D-63	Sequence 63, Appl
51	0.2	8342	5	PCT-US94-04496-63	Sequence 63, Appl
51	0.2	8392	1	US-08-080-255-6	Sequence 6, Appl
51	0.2	8392	1	US-08-465-713-6	Sequence 6, Appl
51	0.2	8392	5	PCT-US93-05857-6	Sequence 6, Appl
50	0.2	10684	3	US-08-618-100B-3	Sequence 3, Appl
50	0.2	152331	3	US-09-128-155-16	Sequence 16, Appl
50	0.2	176373	3	US-09-128-155-17	Sequence 17, Appl
50	0.2	246240	2	US-08-724-394A-20	Sequence 20, Appl
50	0.2	246240	2	US-08-724-394A-21	Sequence 21, Appl
50	0.2	246240	2	US-08-724-394A-22	Sequence 22, Appl
49	0.2	9278	1	US-08-243-542-9	Sequence 9, Appl
49	0.2	9278	1	US-08-477-407-9	Sequence 9, Appl
49	0.2	9278	1	US-08-484-355-9	Sequence 9, Appl
48	0.2	282	1	US-08-133-629-8	Sequence 8, Appl
48	0.2	2957	2	US-08-394-152A-48	Sequence 48, Appl
48	0.2	3017	2	US-08-394-152A-39	Sequence 39, Appl
48	0.2	6340	1	US-08-187-785-3	Sequence 3, Appl
48	0.2	8835	3	US-08-884-324-10	Sequence 10, Appl
48	0.2	20303	1	US-08-370-975B-6	Sequence 6, Appl
48	0.2	26764	1	US-08-370-975B-1	Sequence 1, Appl
47	0.2	28994	3	US-08-884-324-14	Sequence 14, Appl
46	0.2	1664	1	US-08-250-740-34	Sequence 34, Appl
46	0.2	1664	1	US-07-695-472B-3	Sequence 3, Appl
46	0.2	1926	4	US-09-117-250-4	Sequence 4, Appl
46	0.2	2886	2	US-08-687-680-55	Sequence 55, Appl
46	0.2	3195	3	US-08-951-648-5	Sequence 5, Appl
46	0.2	3195	3	US-09-174-437-5	Sequence 5, Appl
46	0.2	4741	1	US-07-695-472B-4	Sequence 4, Appl
46	0.2	4742	1	US-08-250-740-35	Sequence 35, Appl
46	0.2	5408	1	US-08-471-058-20	Sequence 20, Appl
46	0.2	5408	3	US-08-471-057-20	Sequence 20, Appl
46	0.2	246240	2	US-08-724-394A-20	Sequence 20, Appl
46	0.2	246240	2	US-08-724-394A-21	Sequence 21, Appl
46	0.2	246240	2	US-08-724-394A-22	Sequence 22, Appl
45	0.2	1386	2	US-08-687-080-76	Sequence 76, Appl
45	0.2	2373	4	US-08-975-762-45	Sequence 45, Appl
45	0.2	2392	1	US-08-171-382-5	Sequence 5, Appl
45	0.2	2392	1	US-08-309-420-5	Sequence 5, Appl
45	0.2	2392	1	US-08-309-419-5	Sequence 5, Appl
45	0.2	2392	5	PCT-US95-11856-5	Sequence 5, Appl
45	0.2	2392	5	PCT-US95-11878-5	Sequence 5, Appl
45	0.2	3715	4	US-09-085-199B-44	Sequence 44, Appl
45	0.2	5375	3	US-08-757-223-7	Sequence 7, Appl
45	0.2	8353	3	US-08-611-587-1	Sequence 1, Appl
45	0.2	8779	2	US-08-750-703-4	Sequence 4, Appl
45	0.2	11613	1	US-08-484-044-10	Sequence 10, Appl
45	0.2	17041	1	US-08-076-011-1	Sequence 1, Appl
45	0.2	18596	4	US-09-318-448-11	Sequence 11, Appl
45	0.2	35060	3	US-08-814-095-7	Sequence 7, Appl
45	0.2	87350	3	US-08-781-891-79	Sequence 79, Appl
44	0.2	264	1	US-08-222-177A-10	Sequence 10, Appl
44	0.2	289	2	US-08-481-658B-63	Sequence 63, Appl
44	0.2	289	2	US-08-477-504A-63	Sequence 63, Appl
44	0.2	289	2	US-08-486-756A-63	Sequence 63, Appl
44	0.2	289	2	US-08-485-862B-63	Sequence 63, Appl
44	0.2	289	3	US-08-787-739-63	Sequence 63, Appl
44	0.2	289	3	US-08-487-077A-63	Sequence 63, Appl
44	0.2	289	3	US-08-485-049D-63	Sequence 63, Appl
44	0.2	1334	2	US-08-481-658B-44	Sequence 44, Appl
44	0.2	1334	2	US-08-477-504A-44	Sequence 44, Appl
44	0.2	1334	2	US-08-486-756A-44	Sequence 44, Appl
44	0.2	1334	2	US-08-485-862B-44	Sequence 44, Appl
44	0.2	1334	3	US-08-787-739-44	Sequence 44, Appl

Query Match 0.2%; Score 60; DB 3; Length 176373;
Best Local Similarity 100.0%; Pred. No. 6.8e-14;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	11740	tttttttgagacagagtgctcactctgtcgccaggctggagtgcgagtgggcagcatctcgg	11799
Db	143612	tttttttgagacagagtgctcactctgtcgccaggctggagtgcgagtgggcagcatctcgg	143671

```

RESULT      5
US-08-323-443B-1
; Sequence 1, Application US/08323443B
; Patent No. 5654170
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W.
; APPLICANT: LANDES, GREGORY M.
; APPLICANT: BURN, TIMOTHY C.
; APPLICANT: CONNORS, TIMOTHY D.
; APPLICANT: DACKOWSKI, WILLIAM R.
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dardy & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA

```

```
Query Match          0.2%; Score 59; DB 1; Length 31571;
Best Local Similarity 100.0%; Pred. No. 2e-13;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

[illegible]

RESULT 6

```

US-08-658-136-2
; Sequence 2, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5356 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-658-136-2

```

```
Query Match          0.2%; Score 59; DB 3; Length 53526;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 12652 agagacaggggtttctccacgttggtcaggctggtctcaaacctcctgacctcaggtgatc 12710
 |||||
 Db 7254 AGAGACGGGGGTTTCTCCACGTTGGTCAGGCTGGTCTCAACTCCTGACCTCAGGTGATC 7312

```

RESULT      7
US-08-658-136-1
: Sequence 1, Application US/08658136
: Patent No. 6071717
: GENERAL INFORMATION:
: APPLICANT: KLINGER, KATHERINE W
: APPLICANT: LANDES, GREGORY M
: APPLICANT: BURN, TIMOTHY C
: APPLICANT: CONNORS, TIMOTHY D
: APPLICANT: DACKOWSKI, WILLIAM
: APPLICANT: GERMINO, GREGORY
: APPLICANT: QIAN, FENG
: TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
: NUMBER OF SEQUENCES: 58
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: GENZYME CORPORATION
: STREET: ONE MOUNTAIN ROAD
: CITY: FRAMINGHAM

```

; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53577 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-658-136-1

Query Match 0.2%; Score 59; DB 3; Length 53577;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12652 agaacgaggggtttccacggtggtcagctggtctcaaaactcctgacctcaggtgac 12710
|||||
DB 7253 AGAGCGGGGTTCTCCACGTTGGTCAGCGTGGTCTCAAACTCTCGACCTCAGGCGATC 7311

RESULT 8
US-08-909-965C-1/c
; Sequence 1, Application US/08909965C
; Patent No. 5936078
; GENERAL INFORMATION:
; APPLICANT: Kuga Tetsuo
; APPLICANT: Nakagawa Satoshi
; APPLICANT: Sakaki Yoshiyuki
; APPLICANT: Zhao Nanding
; APPLICANT: Hashida Hideji
; TITLE OF INVENTION: NOVEL DNA, NOVEL POLYPEPTIDE
; TITLE OF INVENTION: AND NOVEL ANTIBODY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FITZPATRICK, CELLA, HARPER AND SCINTO
; STREET: 277 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10172-0194
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/909,965C
; FILING DATE: August 12, 1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 322745/95
; APPLICATION NUMBER: PCT/JP96/03630
; FILING DATE: 12-Dec-1996
; ATTORNEY/AGENT INFORMATION:

; NAME: Lawrence S. Perry
; REGISTRATION NUMBER: 31865
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-2400
; TELEFAX: 212-758-2982
; TELEX: 236262
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2688 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: human
; IMMEDIATE SOURCE:
; CLONE: F55
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 358 to 486
; LOCATION: 560 to 799
; LOCATION: 1042 to 1182
; LOCATION: 2105 to 2269
; LOCATION: 2370 to 2462
; IDENTIFICATION METHOD: by experiment
; US-08-909-965C-1

Query Match 0.2%; Score 58; DB 2; Length 2688;
Best Local Similarity 100.0%; Pred. No. 6.2e-13;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17358 ttttttgagcaggtttcactctgttccacgctggagtgcaatggcgatct 17415
|||||
DB 1312 TTTTGGAGAGGAGTTTCACTCTGTGTCGCCAGGCTGGAGTGCATATGGCGTGAICT 1255

RESULT 9
US-08-417-174-1
; Sequence 1, Application US/08417174
; Patent No. 5844075
; GENERAL INFORMATION:
; APPLICANT: KAWAKAMI, YUTAKA; ROSENBERG,
; APPLICANT: STEVEN A.
; TITLE OF INVENTION: MELANOMA ANTIGENS AND
; TITLE OF INVENTION: THEIR USE IN DIAGNOSTIC AND THERAPEUTIC
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 126
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,174
; FILING DATE: 05-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,565
; FILING DATE: 22-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CAROL M. GRUPPI
; REGISTRATION NUMBER: 37,341
; REFERENCE/DOCKET NUMBER: 2026-4124US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800

; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1559
; TYPE: NUCLEOTIDE
; STRANDEDNESS: DOUBLE
; TOPOLOGY: UNKNOWN
; MOLECULE TYPE: CDNA
US-08-417-174-1

Query Match 0.2%; Score 57; DB 2; Length 1559;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12671 gttggtcaggctggtctcaaacctcctgacctcaggatgatctgccgcctcagctccc 12727
|||||
Db 1162 GTTGGTCAGGCTGGTCTCAAACTCCTGACCTCAGGTGATCTGCCGCCCTCAGCCTCC 1218

RESULT 10
US-08-231-565A-1
; Sequence 1, Application US/08231565A
; Patent No. 5874560
; GENERAL INFORMATION:
; APPLICANT: KAWAKAMI, YUTAKA; ROSENBERG,
; APPLICANT: STEVEN A.
; TITLE OF INVENTION: MELANOMA ANTIGENS AND
; TITLE OF INVENTION: THEIR USE IN DIAGNOSTIC AND THERAPEUTIC
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,565A
; FILING DATE: 22-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CAROL M. GRUPPI
; REGISTRATION NUMBER: 37,341
; REFERENCE/DOCKET NUMBER: 2026-4124
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1559
; TYPE: NUCLEOTIDE
; STRANDEDNESS: DOUBLE
; TOPOLOGY: UNKNOWN
; MOLECULE TYPE: CDNA
US-08-231-565A-1

Query Match 0.2%; Score 57; DB 2; Length 1559;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12671 gttggtcaggctggtctcaaacctcctgacctcaggatgatctgccgcctcagctccc 12727
|||||
Db 1162 GTTGGTCAGGCTGGTCTCAAACTCCTGACCTCAGGTGATCTGCCGCCCTCAGCCTCC 1218

RESULT 11
US-09-007-961-1
; Sequence 1, Application US/09007961
; Patent No. 5994523
; GENERAL INFORMATION:
; APPLICANT: KAWAKAMI, YUTAKA; ROSENBERG,
; APPLICANT: STEVEN A.
; TITLE OF INVENTION: MELANOMA ANTIGENS AND
; TITLE OF INVENTION: THEIR USE IN DIAGNOSTIC AND THERAPEUTIC
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/007,961
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/231,565
; FILING DATE: 22-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: CAROL M. GRUPPI
; REGISTRATION NUMBER: 37,341
; REFERENCE/DOCKET NUMBER: 2026-4124
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1559
; TYPE: NUCLEOTIDE
; STRANDEDNESS: DOUBLE
; TOPOLOGY: UNKNOWN
; MOLECULE TYPE: CDNA
US-09-007-961-1

Query Match 0.2%; Score 57; DB 2; Length 1559;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12671 gttggtcaggctggtctcaaacctcctgacctcaggatgatctgccgcctcagctccc 12727
|||||
Db 1162 GTTGGTCAGGCTGGTCTCAAACTCCTGACCTCAGGTGATCTGCCGCCCTCAGCCTCC 1218

RESULT 12
US-08-370-975B-6
; Sequence 6, Application US/08370975B
; Patent No. 5622851
; GENERAL INFORMATION:
; APPLICANT: Maley, Frank
; APPLICANT: Maley, Gladys F.
; APPLICANT: Weiner, Karen X.B.
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester

; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,975B
; FILING DATE: 10-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20894/80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)263-1636
; TELEFAX: (716)263-1600
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 4q35
; US-08-370-975B-6

Query Match 0.2%; Score 55; DB 1; Length 20303;
Best Local Similarity 100.0%; Pred. No. 6.8e-12;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9013 gtttaatttttatttttagtagagatggggtttcaccatgttgccaggctggt 9067
|||||
Db 6385 GTTAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTGGCCAGGCTGTT 6439

RESULT 13
US-08-370-975B-1
; Sequence 1, Application US/08370975B
; Patent No. 5622851
; GENERAL INFORMATION:
; APPLICANT: Maley, Frank
; APPLICANT: Maley, Gladys F.
; APPLICANT: Weiner, Karen X.B.
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,975B
; FILING DATE: 10-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20894/80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)263-1636
; TELEFAX: (716)263-1600

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26764 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 4q35
; US-08-370-975B-1

Query Match 0.2%; Score 55; DB 1; Length 26764;
Best Local Similarity 100.0%; Pred. No. 6.6e-12;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9013 gtttaatttttatttttagtagagatggggtttcaccatgttgccaggctggt 9067
|||||
Db 8348 GTTAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTGGCCAGGCTGTT 8402

RESULT 14
US-08-480-784-20
; Sequence 20, Application US/08480784
; Patent No. 5693473
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,784
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957

REFERENCE/DOCKET NUMBER: 24884-109347
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-480-784-20

Query Match 0.2%; Score 53; DB 1; Length 6769;
Best Local Similarity 100.0%; Pred. No. 4.4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtatttttagtagagatggggttcaccatgttggccaggtcgtt 9067
|||||
Db 5609 TAATTTTGTATTTTAGTAGATGGGTTTCACCATGTTGGCCAGGCTGGT 5661

RESULT 15

US-08-483-553-20
Sequence 20, Application US/08483553
Patent No. 5709999

GENERAL INFORMATION:

APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,553
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/309,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-483-553-20

Query Match 0.2%; Score 53; DB 1; Length 6769;
Best Local Similarity 100.0%; Pred. No. 4.4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtatttttagtagagatggggttcaccatgttggccaggtcgtt 9067
|||||
Db 5609 TAATTTTGTATTTTAGTAGATGGGTTTCACCATGTTGGCCAGGCTGGT 5661

RESULT 16

US-08-487-002-20
Sequence 20, Application US/08487002
Patent No. 5710001

GENERAL INFORMATION:

APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Simard, Jacques
APPLICANT: Emi, Mitsuru
APPLICANT: Nakamura, Yusuke
APPLICANT: Durocher, Francine
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,002
FILING DATE:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994

```
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6769 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
;
US-08-487-002-20

Query Match          0.2%; Score 53; DB 1; Length 6769;
Best Local Similarity 100.0%; Pred. No. 4.4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagatagagatgggtttccaccatgttgccaggtcgtg 9067
      |||||||
Db 5609 TAAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGCT 5661

RESULT 17
US-08-483-554B-20
; Sequence 20, Application US/08483554B
; Patent No. 5747282
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,554B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
```

```
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6769 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
;
US-08-483-554B-20

Query Match          0.2%; Score 53; DB 1; Length 6769;
Best Local Similarity 100.0%; Pred. No. 4.4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagatagagatgggtttccaccatgttgccaggtcgtg 9067
      |||||||
Db 5609 TAAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGCT 5661

RESULT 18
US-08-488-011B-20
; Sequence 20, Application US/08488011B
; Patent No. 5753441
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,011B
; FILING DATE: 07-JUN-1995
```


CLASSIFICATION: 435
PRIOR APPLICATION DATA: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347-09
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-488-011B-20

Query Match 0.2% Score 53; DB 1; Length 6769;
Best Local Similarity 100.0%; Pred. No. 4.4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagtagagatgggtttccaccatgttggccagctggt 9067
|||||
Db 5609 TAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGT 5661

RESULT 19
US-08-850-727-20
Sequence 20, Application US/08850727
Patent No. 6162897
GENERAL INFORMATION:
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavtigian, Sean W.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/850,727
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: US 08/483,554
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-850-727-20

Query Match 0.2% Score 53; DB 4; Length 6769;
Best Local Similarity 100.0%; Pred. No. 4.4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagtagagatgggtttccaccatgttggccagctggt 9067
|||||
Db 5609 TAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGT 5661

RESULT 20
PCT-US95-10202-20
Sequence 20, Application PC/TUS9510202
GENERAL INFORMATION:
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Simard, Jacques
APPLICANT: Emi, Mitsuru
APPLICANT: Nakamura, Yusuke
APPLICANT: Durocher, Francine
TITLE OF INVENTION: In Vivo Mutations and Polymorphisms
TITLE OF INVENTION: in the 17q-Linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10202
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08-308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6769 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; PCT-US95-10202-20

Query Match 0.2%; Score 53; DB 5; Length 6769;
Best Local Similarity 100.0%; Pred. No. 4.4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9015 taattttttagatagagatgggtttccaccatgttggccaggctggt 9067
Db 5609 TAAATTTTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCT 5661

RESULT 21
PCT-US95-10203-20
; Sequence 20, Application PC/TUS9510203
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10203
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08-308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6769 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; PCT-US95-10203-20

Query Match 0.2%; Score 53; DB 5; Length 6769;
Best Local Similarity 100.0%; Pred. No. 4.4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttttagatagagatgggtttccaccatgttggccaggctggt 9067
Db 5609 TAAATTTTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCT 5661

RESULT 22
PCT-US95-10220-20
; Sequence 20, Application PC/TUS9510220
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander

APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eldens, Donna M.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: Method for Diagnosing a
TITLE OF INVENTION: Predisposition for Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10220
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08-308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
PCT-US95-10220-20

Query Match 0.28; Score 53; DB 5; Length 6769;
Best Local Similarity 100.0%; Pred. No. 4.4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagtagagatgggtttccaccatgttggccagctggt 9067
|||||

Db 5609 TAAATTTGTATTTTGTAGTAGAGATGGGTTTCCACCATGTGGCCAGCTGGT 5661

RESULT 23

US-09-318-448-5
Sequence 5, Application US/09318448
Patent No. 6210950
GENERAL INFORMATION:
APPLICANT: Johnson, William G.
APPLICANT: Stenroos, Edward S.
TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
FILE REFERENCE: 601-1-057
CURRENT APPLICATION NUMBER: US/09/318,448
CURRENT FILING DATE: 1999-05-25
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 7720
TYPE: DNA
ORGANISM: Homo sapiens
US-09-318-448-5

Query Match 0.28; Score 53; DB 4; Length 7720;
Best Local Similarity 100.0%; Pred. No. 4.3e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4709 gctaattttgtatttttagtagagacgggtttccaccatgttggccagat 4761
|||||

Db 5960 gctaattttgtatttttagtagagacgggtttccaccatgttggccagat 6012

RESULT 24

US-09-009-217-11/C
Sequence 11, Application US/09009217
Patent No. 6132729
GENERAL INFORMATION:
APPLICANT: Thorpe, Philip E.
APPLICANT: King, Steven W.
APPLICANT: Gao, Boning
TITLE OF INVENTION: COMBINED TISSUE FACTOR AND
TITLE OF INVENTION: CHEMOTHERAPEUTIC METHODS AND COMPOSITIONS FOR COAGULATION
TITLE OF INVENTION: AND TUMOR TREATMENT
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,217
FILING DATE: Concurrently Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/042,427
FILING DATE: 27-MAR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/036,205
FILING DATE: 27-JAN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/035,920
FILING DATE: 22-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hibler, David W.
REGISTRATION NUMBER: 41,071
REFERENCE/DOCKET NUMBER: UTSD:536
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577

; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13865 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-009-217-11

Query Match 0.2%; Score 53; DB 3; Length 13865;
Best Local Similarity 100.0%; Pred. No. 4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtatttttagatggggtttccaccatgttgccagctggt 9067
|||||
DB 8526 TAATTTTGTATTATTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCT 8474

RESULT 25
US-09-009-656-11/c
; Sequence 11, Application US/09009656
; Patent No. 6132730
; GENERAL INFORMATION:
; APPLICANT: Thorpe, Philip E.
; APPLICANT: King, Steven W.
; APPLICANT: Gao, Boming
; TITLE OF INVENTION: COMBINED TISSUE FACTOR AND FACTOR VIIa
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR COAGULATION AND TUMOR
; TITLE OF INVENTION: TREATMENT
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,656
; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/042,427
; FILING DATE: 27-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/036,205
; FILING DATE: 27-JAN-1997

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/035,920
; FILING DATE: 22-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hibler, David W.
; REGISTRATION NUMBER: 41,071
; REFERENCE/DOCKET NUMBER: UTSD:537
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577

; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13865 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-009-656-11

Query Match 0.2%; Score 53; DB 3; Length 13865;
Best Local Similarity 100.0%; Pred. No. 4e-11;

Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9015 taattttgtatttttagatggggtttccaccatgttgccagctggt 9067
|||||
DB 8526 TAATTTTGTATTATTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCT 8474

RESULT 26
US-09-009-913-1
; Sequence 1, Application US/09009913
; Patent No. 6087485
; GENERAL INFORMATION:
; APPLICANT: Axy's Pharmaceuticals, Inc.
; TITLE OF INVENTION: Asthma Related Genes
; NUMBER OF SEQUENCES: 339
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,913
; FILING DATE: 21-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SEQ-4P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3231
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-009-913-1

Query Match 0.2%; Score 53; DB 3; Length 72928;
Best Local Similarity 100.0%; Pred. No. 3.4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtatttttagatggggtttccaccatgttgccagctggt 9067
|||||
DB 55385 TAATTTTGTATTATTAGTAGAGATGGGGTTTCACCAITGTTGGCCAGGCTGCT 55437

RESULT 27
US-08-781-891-79/c
; Sequence 79, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME

; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 87350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-781-891-79

Query Match 0.2%; Score 53; DB 3; Length 87350;
Best Local Similarity 100.0%; Pred. No. 3.3e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtatttagtagagatgggtttccaccatgttgccagctggt 9067
|||||
Db 68060 TAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGT 68008

RESULT 28
US-08-531-927B-9/c
; Sequence 9, Application US/08531927B
; Patent No. 5840491
; GENERAL INFORMATION:
; APPLICANT: Kakizuka, Akira
; TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
; Patent No. 5840491
; TITLE OF INVENTION: Disease Gene and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/531,927B
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP H6-251600
; FILING DATE: 21-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia

; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: ATH95-01A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-531-927B-9

Query Match 0.2%; Score 52; DB 2; Length 807;
Best Local Similarity 100.0%; Pred. No. 1.3e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9016 aattttgtatttagtagagatgggtttccaccatgttgccagctggt 9067
|||||
Db 618 AATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGT 567

RESULT 29
US-08-531-927B-1/c
; Sequence 1, Application US/08531927B
; Patent No. 5840491
; GENERAL INFORMATION:
; APPLICANT: Kakizuka, Akira
; TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
; Patent No. 5840491
; TITLE OF INVENTION: Disease Gene and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/531,927B
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP H6-251600
; FILING DATE: 21-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: ATH95-01A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1776 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36..1115
; US-08-531-927B-1

Query Match 0.2%; Score 52; DB 2; Length 1776;

```
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9016 aattttgtatttttagtagagtggtttccaccatgttgccaggctggt 9067
|||||
Db 1605 AATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGCT 1554

RESULT 30
US-09-041-886-12/c
; Sequence 12, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroo
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041.886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1776 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36..1116
US-09-041-886-12

Query Match 0.2%; Score 52; DB 4; Length 1776;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9016 aattttgtatttttagtagagtggtttccaccatgttgccaggctggt 9067
|||||
Db 1605 AATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGCT 1554

RESULT 31
US-07-906-871-15/c
; Sequence 15, Application US/07906871
; Patent No. 5340739
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Avraham, Shalom
; TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES
; THEREOF
```

```
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Avenue, N.W., Suite 300
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/906.871
; FILING DATE: 19920103
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/816.289
; FILING DATE: 03 JAN 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/635.544
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US89/03051
; FILING DATE: 13-JUL-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/224.035
; FILING DATE: 13-JUL-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbala, Michele A
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 0627.2830004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)833-7533
; TELEFAX: (202)833-8716
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17327 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: exon
; LOCATION: 621..753
; FEATURE:
; NAME/KEY: intron
; LOCATION: 754..9596
; FEATURE:
; NAME/KEY: exon
; LOCATION: 9597..9744
; FEATURE:
; NAME/KEY: intron
; LOCATION: 9745..16396
; FEATURE:
; NAME/KEY: exon
; LOCATION: 16397..17327
; US-07-906-871-15

Query Match 0.2%; Score 52; DB 1; Length 17327;
Best Local Similarity 100.0%; Pred. No. 9.4e-11;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8932 cctccacctccagggttcaagtattctctgcctcagctcccaagtagct 8983
|||||
Db 5233 CCTCCACCTCCAGGTTCAAGTGATTCTCTGCCTCAGGCTCCCAAGTAGCT 5182

RESULT 32
US-08-351-149-4/c
; Sequence 4, Application US/08351149
```

```
; Patent No. 5629283
; GENERAL INFORMATION:
; APPLICANT: Nicola, Nicos A.
; APPLICANT: Gough, Nicholas M.
; APPLICANT: Gearing, David P.
; APPLICANT: Metcalf, Donald
; APPLICANT: King, Julie Ann
; TITLE OF INVENTION: Improvements in Granulocyte-Macrophage
; TITLE OF INVENTION: Colony-Stimulating Factor Receptor and Derivatives Thereof
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/351,149
; FILING DATE: 23-NOV-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Fox, Samuel L.
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 1256.0030001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1808 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..90
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 150..1349
; US-08-351-149-4

Query Match 0.2%; Score 51; DB 1; Length 1808;
Best Local Similarity 100.0%; Pred. No. 2.9e-10;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10565 tgatccgcctgcttggcctcccaaaagtgtgggattacagcgctgagcca 10615
|||||
Db 1540 TGATCCGCCTGCTGGCTGCCCTCCCAAGTGTGGGATTACAGCGCTGAGCCA 1490

RESULT 33
US-08-384-828-4/c
; Sequence 4, Application US/08384828
; Patent No. 5726036
; GENERAL INFORMATION:
; APPLICANT: Nicola, Nicos A.
; APPLICANT: Gough, Nicholas M.
; APPLICANT: Gearing, David P.
; APPLICANT: Metcalf, Donald
; APPLICANT: King, Julie Ann
; TITLE OF INVENTION: Improvements in Granulocyte-Macrophage
; TITLE OF INVENTION: Colony-Stimulating Factor Receptor and Derivatives Thereof
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX
; STREET: 1100 New York Ave., NW
```

```
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/384,828
; FILING DATE: 07-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/351,149
; FILING DATE: 23-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fox, Samuel L.
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 1256.0030001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1808 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..90
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 150..1349
; US-08-384-828-4

Query Match 0.2%; Score 51; DB 1; Length 1808;
Best Local Similarity 100.0%; Pred. No. 2.9e-10;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10565 tgatccgcctgcttggcctcccaaaagtgtgggattacagcgctgagcca 10615
|||||
Db 1540 TGATCCGCCTGCTGGCTGCCCTCCCAAGTGTGGGATTACAGCGCTGAGCCA 1490

RESULT 34
US-08-895-474-4/c
; Sequence 4, Application US/08895474
; Patent No. 6136957
; GENERAL INFORMATION:
; APPLICANT: Nicola, Nicos A.
; APPLICANT: Gough, Nicholas M.
; APPLICANT: Gearing, David P.
; APPLICANT: Metcalf, Donald
; APPLICANT: King, Julie Ann
; TITLE OF INVENTION: Improvements in Granulocyte-Macrophage
; TITLE OF INVENTION: Colony-Stimulating Factor Receptor and Derivatives Thereof
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Ave., NW, Ste. 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/895,474
FILING DATE: 16-JUL-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Fox, Samuel L.
REGISTRATION NUMBER: 30,353
REFERENCE/DOCKET NUMBER: 1256.0030003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1808 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
FEATURE:
NAME/KEY: CDS
LOCATION: 25...90
FEATURE:
NAME/KEY: CDS
LOCATION: 150..1349
US-08-895-474-4

Query Match 0.2%; Score 51; DB 3; Length 1808;
Best Local Similarity 100.0%; Pred. No. 2.9e-10;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10565 tgatccgctgctggctcccaagtgctgggattacagcgtagccca 10615
|||||
Db 1540 TGATCCGCTGCTGGCTCCCAAGTGTGGATTACAGCGTAGGCCA 1490

RESULT 35

US-08-886-152-4/c
Sequence 4, Application US/08886152
Patent No. 5880273
GENERAL INFORMATION:
APPLICANT: ADACHI, HIDEKI
APPLICANT: TSUJIMOTO, MASAFUMI
APPLICANT: ARAI, HIROYUKI
APPLICANT: INOUE, KEIZO
TITLE OF INVENTION: PLATELET ACTIVATING FACTOR
TITLE OF INVENTION: ACETYLHYDROLASE, AND GENE THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/886,152
FILING DATE: 30-JUN-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-188369
FILING DATE: 28-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2292-041-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2559 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE: HUMAN
FEATURE:
NAME/KEY: CDS
LOCATION: 216..1392
US-08-886-152-4

Query Match 0.2%; Score 51; DB 2; Length 2559;
Best Local Similarity 100.0%; Pred. No. 2.8e-10;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4709 gctaattttttagtagagacggggtttcaccatgttgccagg 4759
|||||
Db 2415 GCTAATTTTGTATTTTAGTAGAGACGGGTTTCACCATGTGCGCAGG 2365

RESULT 36

US-08-545-860D-63/c
Sequence 63, Application US/08545860D
Patent No. 6040140
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaani, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
ADDRESSEE: No. 6040140ris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,860D
FILING DATE: 07-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE: 22-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10930
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/327,392
FILING DATE: 19-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/320,559
FILING DATE: 11-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,443
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,094
FILING DATE: 30-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,839
FILING DATE: 27-MAY-1992

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/805,093
;; FILING DATE: 11-DEC-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: DeLuca Esq., Mark
;; REGISTRATION NUMBER: 33,229
;; REFERENCE/DOCKET NUMBER: TJU-1262
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 568-3100
;; TELEFAX: (215) 568-3439
;; INFORMATION FOR SEQ ID NO: 63:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 8342 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 2..265
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 595..666
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 2353..2484
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3032..3145
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 6788..6934
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 7967..8062
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 8304..8342
;; US-08-545-860D-63

Query Match 0.2%; Score 51; DB 3; Length 8342;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10565 tgatccgcctgcttgccctcccaaaagtgtgggattacagcgtagccca 10615
|||||
Db 4035 TGATCCGCCTGCTTGCCCTCCCAAAAGTGTGGGATTACAGCGTAGGCCA 3985

RESULT 37
PCT-US94-04496-63/c
; Sequence 63, Application PC/TUS9404496
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaan, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US94/04496
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: DeLuca Esq., Mark
;; REGISTRATION NUMBER: 33,229
;; REFERENCE/DOCKET NUMBER: TJU-1242
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 568-3100
;; TELEFAX: (215) 568-3439
;; INFORMATION FOR SEQ ID NO: 63:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 8342 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 2..265
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 595..666
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 2353..2484
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3032..3145
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 6788..6934
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 7967..8062
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 8304..8342
;; PCT-US94-04496-63

Query Match 0.2%; Score 51; DB 5; Length 8342;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10565 tgatccgcctgcttgccctcccaaaagtgtgggattacagcgtagccca 10615
|||||
Db 4035 TGATCCGCCTGCTTGCCCTCCCAAAAGTGTGGGATTACAGCGTAGGCCA 3985

RESULT 38
US-08-080-255-6/c
; Sequence 6, Application US/08080255
; Patent No. 5487970
; GENERAL INFORMATION:
; APPLICANT: Rowley, Janet D.
; APPLICANT: Diaz, Manuel O.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND
; TITLE OF INVENTION: TRANSLOCATIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/080,255
FILING DATE: 19930617
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: ARCD:072/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 320-7200
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 8392 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-080-255-6

Query Match 0.2%; Score 51; DB 1; Length 8392;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10565 tgatccgcctgcttggcctcccaagtctgggattacagcgtagccca 10615
|||||
Db 4086 TGATCCGCTGCTGGCCTCCCAAAGTCTGGGATTACAGCGTAGGCCA 4036

RESULT 39
US-08-465-713-6/c
Sequence 6, Application US/08465713
Patent No. 6121419
GENERAL INFORMATION:
APPLICANT: Rowley, Janet D.
APPLICANT: Diaz, Manuel O.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND
TITLE OF INVENTION: TRANSLOCATIONS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,713
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/080,255
FILING DATE: 17 JUNE 1993
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: ARCD:072/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 320-7200
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 8392 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-465-713-6
Query Match 0.2%; Score 51; DB 3; Length 8392;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10565 tgatccgcctgcttggcctcccaagtctgggattacagcgtagccca 10615
|||||
Db 4086 TGATCCGCTGCTGGCCTCCCAAAGTCTGGGATTACAGCGTAGGCCA 4036

RESULT 40
PCT-US93-05857-6/c
Sequence 6, Application PC/TUS9305857
GENERAL INFORMATION:
APPLICANT: Board of Regents
APPLICANT: The University of Texas System
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
TITLE OF INVENTION: GENE REARRANGEMENTS AND TRANSLOCATIONS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/05857
FILING DATE: 19930617
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/900,689
FILING DATE: 17/06/92
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: ARCD:072/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 320-7200
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 8392 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US93-05857-6

Query Match 0.2%; Score 51; DB 5; Length 8392;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10565 tgatccgcctgcttggcctcccaagtctgggattacagcgtagccca 10615
|||||
Db 4086 TGATCCGCTGCTGGCCTCCCAAAGTCTGGGATTACAGCGTAGGCCA 4036

RESULT 41

US-08-618-100B-3/c
; Sequence 3, Application US/08618100B
; Patent No. 6068976
; GENERAL INFORMATION:
; APPLICANT: Briggs, Michael R.
; APPLICANT: Auwerx, Johan
; APPLICANT: de Vos, Piet
; APPLICANT: Staelens, Bart
; APPLICANT: Croston, Glenn E.
; APPLICANT: Miller, Stephen G.
; TITLE OF INVENTION: MODULATORS OF OB GENE AND
; TITLE OF INVENTION: SCREENING METHODS THEREFOR
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/618,100B
; FILING DATE: March 19, 1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/558,588
; FILING DATE: October 30, 1995
; APPLICATION NUMBER: 08/510,584
; FILING DATE: August 2, 1995
; APPLICATION NUMBER: 08/418,096
; FILING DATE: April 5, 1995
; APPLICATION NUMBER: 08/408,584
; FILING DATE: March 20, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 219/075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10684 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; DESCRIPTION: Sequence between exon 1 and exon 2
; Patent No. 6068976
US-08-618-100B-3

Query Match 0.2%; Score 50; DB 3; Length 10684;
Best Local Similarity 100.0%; Pred. No. 5.7e-10;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9018 tttttgtatttttagtagagatgggtttccaccattgttgccaggctggt 9067
|||||
Db 6836 TTTTGTATTATTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGT 6787

RESULT 42
US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654

; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 0.2%; Score 50; DB 3; Length 152331;
Best Local Similarity 100.0%; Pred. No. 4.2e-10;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9018 tttttgtatttttagtagagatgggtttccaccattgttgccaggctggt 9067
|||||
Db 58280 TTTTGTATTATTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGT 58231

RESULT 43
US-09-128-155-17/c
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Query Match 0.2%; Score 50; DB 3; Length 176373;
Best Local Similarity 100.0%; Pred. No. 4.2e-10;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9018 tttttgtatttttagtagagatgggtttccaccattgttgccaggctggt 9067
|||||
Db 127160 TTTTGTATTATTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGT 127111

RESULT 44
US-08-724-394A-20
; Sequence 20, Application US/08724394A

```
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
;
US-08-724-394A-20

Query Match 0.2%; Score 50; DB 2; Length 246240;
Best Local Similarity 100.0%; Pred. No. 4e-10;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12671 gttggtcagggtgtctcaaaactcctgacctcaggtgatctgccgcctc 12720
|||||
Db 70232 GTTGTGAGGCTGGTCTCAAACTCCTGACCTCAGGTGATCTGCCGCCTC 70281

RESULT 45
US-08-724-394A-21
; Sequence 21, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
;
US-08-724-394A-21

Query Match 0.2%; Score 50; DB 2; Length 246240;
Best Local Similarity 100.0%; Pred. No. 4e-10;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12671 gttggtcagggtgtctcaaaactcctgacctcaggtgatctgccgcctc 12720
|||||
Db 70232 GTTGTGAGGCTGGTCTCAAACTCCTGACCTCAGGTGATCTGCCGCCTC 70281

RESULT 46
US-08-724-394A-22
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H. CONTIG"
; US-08-724-394A-22

Query Match 0.2%; Score 50; DB 2; Length 246240;
Best Local Similarity 100.0%; Pred. No. 4e-10;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12671 gttggtcaggctgtctcaactctgcacctcaggtgatctgccgcctc 12720
|||||
Db 70232 GTTGGTCAGGCTGTCTCAAACTCTGACCTCAGGTGATCTGCCGCCTC 70281

RESULT 47
US-08-243-542-9/c
; Sequence 9, Application US/08243542
; Patent No. 5552526
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: EMI, MITSURU
; TITLE OF INVENTION: MDC PROTEINS AND DNAs
; TITLE OF INVENTION: ENCODING THE SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ELYNN, THIEL, BOUTELL & TANIS P.C.
; STREET: 2026 Rambling Road
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49008-1699
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
; COMPUTER: IBM PC/XT/AT Compatible
; OPERATING SYSTEM: MS-DOS 5.0
; SOFTWARE: Wordperfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/243,542
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-136602
; FILING DATE: 14 MAY 1993
; APPLICATION NUMBER: JP 5-257455
; FILING DATE: 22 SEPTEMBER 1993
; APPLICATION NUMBER: JP 6-49904
; FILING DATE: 23 FEBRUARY 1994
; APPLICATION NUMBER: JP 6-73328
; FILING DATE: 12 APRIL 1994
; APPLICATION NUMBER: JP 6-84470
; FILING DATE: 22 APRIL 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Terryence F. Chapman

; REGISTRATION NUMBER: 32 549
; REFERENCE/DOCKET NUMBER: Furuya Case 1313
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (616) 381-1156
; TELEFAX: (616) 381-5465
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9278 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; LIBRARY: human DNA cosmid library
; FEATURE:
; NAME/KEY: exon 1
; LOCATION: 28..44
; FEATURE:
; NAME/KEY: exon 2
; LOCATION: 308..374
; FEATURE:
; NAME/KEY: exon 3
; LOCATION: 909..994
; FEATURE:
; NAME/KEY: exon 4
; LOCATION: 1081..1156
; FEATURE:
; NAME/KEY: exon 5
; LOCATION: 1591..1657
; FEATURE:
; NAME/KEY: exon 6
; LOCATION: 1725..1792
; FEATURE:
; NAME/KEY: exon 7
; LOCATION: 2182..2256
; FEATURE:
; NAME/KEY: exon 8
; LOCATION: 2339..2410
; FEATURE:
; NAME/KEY: exon 9
; LOCATION: 2588..2754
; FEATURE:
; NAME/KEY: exon 10
; LOCATION: 3248..3332
; FEATURE:
; NAME/KEY: exon 11
; LOCATION: 3445..3535
; FEATURE:
; NAME/KEY: exon 12
; LOCATION: 3645..3696
; FEATURE:
; NAME/KEY: exon 13
; LOCATION: 4014..4113
; FEATURE:
; NAME/KEY: exon 14
; LOCATION: 4196..4267
; FEATURE:
; NAME/KEY: exon 15
; LOCATION: 4386..4478
; FEATURE:
; NAME/KEY: exon 16
; LOCATION: 4920..5000
; FEATURE:
; NAME/KEY: exon 17
; LOCATION: 5347..5397
; FEATURE:
; NAME/KEY: exon 18
; LOCATION: 5501..5564
; FEATURE:
; NAME/KEY: exon 19
; LOCATION: 5767..5866
```

FEATURE: ;
NAME/KEY: exon 20 ;
LOCATION: 6073..6202 ;
FEATURE: ;
NAME/KEY: exon 21 ;
LOCATION: 6300..6468 ;
FEATURE: ;
NAME/KEY: exon 22 ;
LOCATION: 6557..6671 ;
FEATURE: ;
NAME/KEY: exon 23 ;
LOCATION: 6736..6846 ;
FEATURE: ;
NAME/KEY: exon 24 ;
LOCATION: 7829..7846 ;
FEATURE: ;
NAME/KEY: exon 25 ;
LOCATION: 8165..9038 ;
US-08-243-542-9 ;

Query Match 0.28; Score 49; DB 1; Length 9278;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10565 tgatccgctgcttggcctcccaagtgcctgggattacaggcgtgagc 10613
|||||
Db 7075 TGATCCGCTGCCCTGGCTCCCAAGTCTGGGATTACAGCGTGAGC 7027

RESULT 48

US-08-477-407-9/c
Sequence 9, Application US/08477407
Patent No. 5631351

GENERAL INFORMATION:

APPLICANT: NAKAMURA, YUSUKE
APPLICANT: EMI, MITSURU
TITLE OF INVENTION: MDC PROTEINS AND DNAS
TITLE OF INVENTION: ENCODING THE SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLYNN, THIEL, BOUTELL & TANIS P.C.
STREET: 2026 Rambling Road
CITY: Kalamazoo
STATE: Michigan
COUNTRY: USA
ZIP: 49008-1699

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
COMPUTER: IBM PC/XT/AT Compatible
OPERATING SYSTEM: MS-DOS 5.0
SOFTWARE: WordPerfect 5.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,407
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/243,542
FILING DATE: 13-MAY-1994
APPLICATION NUMBER: JP 5-136602
FILING DATE: 14 MAY 1993
APPLICATION NUMBER: JP 5-257455
FILING DATE: 22 SEPTEMBER 1993
APPLICATION NUMBER: JP 6-49904
FILING DATE: 23 FEBRUARY 1994
APPLICATION NUMBER: JP 6-73328
FILING DATE: 12 APRIL 1994
APPLICATION NUMBER: JP 6-84470
FILING DATE: 22 APRIL 1994

ATTORNEY/AGENT INFORMATION:
NAME: Terryence F. Chapman
REGISTRATION NUMBER: 32 549
REFERENCE/DOCKET NUMBER: Furuya Case 1313

TELECOMMUNICATION INFORMATION:
TELEPHONE: (616) 381-1156
TELEFAX: (616) 381-5465
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 9278 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
LIBRARY: human DNA cosmid library
FEATURE:
NAME/KEY: exon 1
LOCATION: 28..44
FEATURE:
NAME/KEY: exon 2
LOCATION: 308..374
FEATURE:
NAME/KEY: exon 3
LOCATION: 909..994
FEATURE:
NAME/KEY: exon 4
LOCATION: 1081..1156
FEATURE:
NAME/KEY: exon 5
LOCATION: 1591..1657
FEATURE:
NAME/KEY: exon 6
LOCATION: 1725..1792
FEATURE:
NAME/KEY: exon 7
LOCATION: 2182..2256
FEATURE:
NAME/KEY: exon 8
LOCATION: 2339..2410
FEATURE:
NAME/KEY: exon 9
LOCATION: 2588..2754
FEATURE:
NAME/KEY: exon 10
LOCATION: 3248..3332
FEATURE:
NAME/KEY: exon 11
LOCATION: 3445..3535
FEATURE:
NAME/KEY: exon 12
LOCATION: 3645..3696
FEATURE:
NAME/KEY: exon 13
LOCATION: 4014..4113
FEATURE:
NAME/KEY: exon 14
LOCATION: 4196..4267
FEATURE:
NAME/KEY: exon 15
LOCATION: 4386..4478
FEATURE:
NAME/KEY: exon 16
LOCATION: 4920..5000
FEATURE:
NAME/KEY: exon 17
LOCATION: 5347..5397
FEATURE:
NAME/KEY: exon 18
LOCATION: 5501..5564
FEATURE:
NAME/KEY: exon 19
LOCATION: 5767..5866
FEATURE:
NAME/KEY: exon 20 ;

LOCATION: 6073..6202
FEATURE:
NAME/KEY: exon 21
LOCATION: 6300..6468
FEATURE:
NAME/KEY: exon 22
LOCATION: 6557..6671
FEATURE:
NAME/KEY: exon 23
LOCATION: 6756..6846
FEATURE:
NAME/KEY: exon 24
LOCATION: 7829..7846
FEATURE:
NAME/KEY: exon 25
LOCATION: 8165..9038
US-08-477-407-9

Query Match 0.2%; Score 49; DB 1; Length 9278;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10565 tgatccgcctgccttgccctcccaagtctgggtacaggtgagc 10613
|||||
Db 7075 TGATCCGCTGCTGGCTGCCCTCCCAAGTGTGGATTACAGCGGTGAGC 7027

RESULT 49

US-08-484-355-9/c
Sequence 9, Application US/08484355
Patent No. 5705341
GENERAL INFORMATION:
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: EMI, MITSURU
TITLE OF INVENTION: MDC PROTEINS AND DNAs
TITLE OF INVENTION: ENCODING THE SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: ELYNN, THIEL, BOUTELL & TANIS P.C.
STREET: 2026 Rambling Road
CITY: Kalamazoo
STATE: Michigan
COUNTRY: USA
ZIP: 49008-1699
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
COMPUTER: IBM PC/XT/AT Compatible
OPERATING SYSTEM: MS-DOS 5.0
SOFTWARE: WordPerfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,355
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/243,542
FILING DATE: 13-MAY-1994
APPLICATION NUMBER: JP 5-136602
FILING DATE: 14 MAY 1993
APPLICATION NUMBER: JP 5-257455
FILING DATE: 22 SEPTEMBER 1993
APPLICATION NUMBER: JP 6-49904
FILING DATE: 23 FEBRUARY 1994
APPLICATION NUMBER: JP 6-73328
FILING DATE: 12 APRIL 1994
APPLICATION NUMBER: JP 6-84470
FILING DATE: 22 APRIL 1994
ATTORNEY/AGENT INFORMATION:
NAME: Terrence F. Chapman
REGISTRATION NUMBER: 32 549
REFERENCE/DOCKET NUMBER: Furuya Case 1313
TELECOMMUNICATION INFORMATION:
TELEPHONE: (616) 381-1156

TELEFAX: (616) 381-5465
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 9278 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
LIBRARY: human DNA cosmid library
FEATURE:
NAME/KEY: exon 1
LOCATION: 28..44
FEATURE:
NAME/KEY: exon 2
LOCATION: 308..374
FEATURE:
NAME/KEY: exon 3
LOCATION: 909..994
FEATURE:
NAME/KEY: exon 4
LOCATION: 1081..1156
FEATURE:
NAME/KEY: exon 5
LOCATION: 1591..1657
FEATURE:
NAME/KEY: exon 6
LOCATION: 1725..1792
FEATURE:
NAME/KEY: exon 7
LOCATION: 2182..2256
FEATURE:
NAME/KEY: exon 8
LOCATION: 2339..2410
FEATURE:
NAME/KEY: exon 9
LOCATION: 2588..2754
FEATURE:
NAME/KEY: exon 10
LOCATION: 3248..3332
FEATURE:
NAME/KEY: exon 11
LOCATION: 3445..3535
FEATURE:
NAME/KEY: exon 12
LOCATION: 3645..3696
FEATURE:
NAME/KEY: exon 13
LOCATION: 4014..4113
FEATURE:
NAME/KEY: exon 14
LOCATION: 4196..4267
FEATURE:
NAME/KEY: exon 15
LOCATION: 4386..4478
FEATURE:
NAME/KEY: exon 16
LOCATION: 4920..5000
FEATURE:
NAME/KEY: exon 17
LOCATION: 5347..5397
FEATURE:
NAME/KEY: exon 18
LOCATION: 5501..5564
FEATURE:
NAME/KEY: exon 19
LOCATION: 5767..5866
FEATURE:
NAME/KEY: exon 20
LOCATION: 6073..6202
FEATURE:

; NAME/KEY: exon 21
; LOCATION: 6300..6468
; FEATURE:
; NAME/KEY: exon 22
; LOCATION: 6557..6671
; FEATURE:
; NAME/KEY: exon 23
; LOCATION: 6756..6846
; FEATURE:
; NAME/KEY: exon 24
; LOCATION: 7829..7846
; FEATURE:
; NAME/KEY: exon 25
; LOCATION: 8165..9038
; US-08-484-355-9

Query Match 0.2%; Score 49; DB 1; Length 9278;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10565 tgatccgctgcttgctggctccaaagtctgggattacagcggtgagc 10613
|||||
Db 7075 TGATCCGCTGCTTGGGCTCCCAAGTCTGGGATTACAGCGGTGAGC 7027

RESULT 50
US-08-133-629-8/c
; Sequence 8, Application US/08133629
; Patent No. 5597694
; GENERAL INFORMATION:
; APPLICANT: Munroe, David J.
; APPLICANT: Housman, David E.
; TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/133,629
; FILING DATE: 07-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Greer, Helen
; REGISTRATION NUMBER: 36,816
; REFERENCE/DOCKET NUMBER: M0828/7001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX: 92-1742 EZEKIEL
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-133-629-8

Query Match 0.2%; Score 48; DB 1; Length 282;
Best Local Similarity 100.0%; Pred. No. 4.9e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11137 tttagtagagacgggtttaccattgtggccaggtgtgcttgaa 11184

Db 125 TTTTtagtagagacgggtttaccattgtggccaggtgtgcttgaa 78
|||||
RESULT 51
US-08-394-152A-48/c
; Sequence 48, Application US/08394152A
; Patent No. 5935818
; GENERAL INFORMATION:
; APPLICANT: Israeli, Ron S.
; APPLICANT: Heston, Warren D.W.
; APPLICANT: Fair, William R.
; TITLE OF INVENTION: PROSTATE-SPECIFIC MEMBRANE ANTIGEN AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM 330 466 DX2
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/394,152A
; FILING DATE: 24-FEB-95
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41426-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2957 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapien
; TISSUE TYPE: Carcinoma
; IMMEDIATE SOURCE:
; CLONE: Prostate Specific Membrane Antigen
; US-08-394-152A-48
Query Match 0.2%; Score 48; DB 2; Length 2957;
Best Local Similarity 100.0%; Pred. No. 3.7e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9015 taattttgtatttttagtagagatgggtttccaccattgtggccagg 9062
|||||
Db 986 TAAATTTTGTATTTTtagtagagatgggtttccaccattgtggccagg 939
|||||
RESULT 52
US-08-394-152A-39
; Sequence 39, Application US/08394152A
; Patent No. 5935818
; GENERAL INFORMATION:
; APPLICANT: Israeli, Ron S.
; APPLICANT: Heston, Warren D.W.
; APPLICANT: Fair, William R.
; TITLE OF INVENTION: PROSTATE-SPECIFIC MEMBRANE ANTIGEN AND

;; TITLE OF INVENTION: USES THEREOF
;; NUMBER OF SEQUENCES: 48
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Cooper & Dunham LLP
;; STREET: 1185 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: United States of America
;; ZIP: 10036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM 330 466 DX2
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/394,152A
;; FILING DATE: 24-FEB-95
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: White, John P.
;; REGISTRATION NUMBER: 28,678
;; REFERENCE/DOCKET NUMBER: 41426-B
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 278-0400
;; TELEFAX: (212) 391-0525
;; INFORMATION FOR SEQ ID NO: 39:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3017 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Homo Sapien
;; TISSUE TYPE: Carcinoma
;; IMMEDIATE SOURCE:
;; CLONE: Prostate Specific Membrane Antigen
;; US-08-394-152A-39

Query Match 0.2%; Score 48; DB 2; Length 3017;
Best Local Similarity 100.0%; Pred. No. 3.7e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtatttttagtagagatgggtttaccacatgttggccagg 9062
|||||
Db 2032 TAATTTTGTATTTTATAGATGGGTTTACCACATGTGGCAGG 2079

RESULT 53
US-08-187-785-3/c
; Sequence 3, Application US/08187785
; Patent No. 5756476
; GENERAL INFORMATION:
; APPLICANT: Epstein, Stephen
; APPLICANT: Unger, Ellis
; APPLICANT: Speir, Edith
; TITLE OF INVENTION: Inhibition of No. 5756476-Transformed Cell
; TITLE OF INVENTION: Proliferation Using Anti-Sense Oligonucleotides
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson, and Bear
; STREET: 620 Newport Center Dr. Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/187,785
;; FILING DATE:
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/07/821,415
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Altman, Daniel E.
;; REGISTRATION NUMBER: 34,115
;; REFERENCE/DOCKET NUMBER: NIH001.001A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 714-760-0404
;; TELEFAX: 714-760-9502
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 6340 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: Human
;; IMMEDIATE SOURCE:
;; CLONE: PCNA Genomic Clone
;; US-08-187-785-3

Query Match 0.2%; Score 48; DB 1; Length 6340;
Best Local Similarity 100.0%; Pred. No. 3.4e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11137 ttttagtagacgggtttcaccatgttggccaggtgtgtctttaa 11184
|||||
Db 230 TTTTGTAGACGAGGGTTTCCACCATGTGGCCAGGCTGCTTGAA 183

RESULT 54
US-08-884-324-10/c
; Sequence 10, Application US/08884324
; Patent No. 6060283
; GENERAL INFORMATION:
; APPLICANT: Takanori OKURA
; APPLICANT: Kakuji TORIGOE
; APPLICANT: Masahi KURIMOTO
; TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE
; TITLE OF INVENTION: OF INDUCING THE PRODUCTION OF INTERFERON-
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/884,324
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 185,305/96
; FILING DATE: 27-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: OKURA-1
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 8835 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: human
TISSUE TYPE: placenta
FEATURE:
NAME/KEY: intron
LOCATION: 1..8835
IDENTIFICATION METHOD: E
US-08-884-324-10

Query Match 0.2%; Score 48; DB 3; Length 8835;
Best Local Similarity 100.0%; Pred. No. 3.3e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12672 ttggtcaggtcgtctcaaacctcctgacctcaggtgatctgcccgcct 12719
|||||
Db 5709 TTGTCAGGCTGGCTCAAACTCTGACCTCAGGTGATCTGCCCGCCT 5662

RESULT 55

US-08-370-975B-6/c
Sequence 6, Application US/08370975B
Patent No. 5622851
GENERAL INFORMATION:
APPLICANT: Maley, Frank
APPLICANT: Maley, Gladys F.
APPLICANT: Weiner, Karen X.B.
TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/370,975B
FILING DATE: 10-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Timian, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 20894/80
TELEPHONE: (716)263-1636
TELEFAX: (716)263-1600
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 20303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 4q35
US-08-370-975B-6

Query Match 0.2%; Score 48; DB 1; Length 20303;
Best Local Similarity 100.0%; Pred. No. 3e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9015 taattttgtatttttagtagagatgggggtttccaccatgttgccagg 9062
|||||
Db 9428 TAAATTTGTATTTTACTAGAGATGGGGTTTCACCATGTTGGCCAGG 9381

RESULT 56

US-08-370-975B-1/c
Sequence 1, Application US/08370975B
Patent No. 5622851
GENERAL INFORMATION:
APPLICANT: Maley, Frank
APPLICANT: Maley, Gladys F.
APPLICANT: Weiner, Karen X.B.
TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/370,975B
FILING DATE: 10-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Timian, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 20894/80
TELEPHONE: (716)263-1636
TELEFAX: (716)263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 26764 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 4q35
US-08-370-975B-1

Query Match 0.2%; Score 48; DB 1; Length 26764;
Best Local Similarity 100.0%; Pred. No. 2.9e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagtagagatgggggtttccaccatgttgccagg 9062
|||||
Db 11391 TAAATTTGTATTTTACTAGAGATGGGGTTTCACCATGTTGGCCAGG 11344

RESULT 57

US-08-884-324-14/c
Sequence 14, Application us/08884324
Patent No. 6060283
GENERAL INFORMATION:
APPLICANT: Takanori OKURA
APPLICANT: Kakuji TORIGOE
APPLICANT: Masahi KURIMOTO

;; TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE
;; OF INDUCING THE PRODUCTION OF INTERFERON-
;; NUMBER OF INVENTIONS: 35
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: BROWDY AND NEIMARK
;; STREET: 419 Seventh Street, N.W., Suite 300
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20004
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/884,324
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 185,305/96
;; FILING DATE: 27-JUN-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BROWDY, Roger L.
;; REGISTRATION NUMBER: 25,618
;; REFERENCE/DOCKET NUMBER: OKURA=1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-628-5197
;; TELEFAX: 202-737-3528
;; INFORMATION FOR SEQ ID NO: 14:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 28994 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: Genomic DNA
;; ORGANISM: human
;; TISSUE TYPE: placenta
;; FEATURE:
;; NAME/KEY: 5'UTR
;; LOCATION: 1..15606
;; IDENTIFICATION METHOD: E
;; NAME/KEY: leader peptide
;; LOCATION: 15607..15685
;; IDENTIFICATION METHOD: S
;; NAME/KEY: intron
;; LOCATION: 15686..17056
;; IDENTIFICATION METHOD: E
;; NAME/KEY: leader peptide
;; LOCATION: 17057..17068
;; IDENTIFICATION METHOD: S
;; NAME/KEY: intron
;; LOCATION: 17069..20451
;; IDENTIFICATION METHOD: E
;; NAME/KEY: leader peptide
;; LOCATION: 20452..20468
;; IDENTIFICATION METHOD: S
;; NAME/KEY: mat peptide
;; LOCATION: 20469..20586
;; IDENTIFICATION METHOD: S
;; NAME/KEY: intron
;; LOCATION: 20587..21920
;; IDENTIFICATION METHOD: E
;; NAME/KEY: mat peptide
;; LOCATION: 21921..22054
;; IDENTIFICATION METHOD: S
;; NAME/KEY: intron
;; LOCATION: 22055..26827
;; IDENTIFICATION METHOD: E
;; NAME/KEY: mat peptide
;; LOCATION: 26828..27046
;; IDENTIFICATION METHOD: S

;; NAME/KEY: 3'UTR
;; LOCATION: 27047..28994
;; IDENTIFICATION METHOD: E
;; US-08-884-324-14
Query Match 0.2%; Score 48; DB 3; Length 28994;
Best Local Similarity 100.0%; Pred. No. 2.9e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12672 ttggctcagggtggtctcaaaactctgacctcaggtgatctgcccgcct 12719
|||||
Db 12472 TTGCTCAGGCTGCTCAAACTCCTGACCTCAGGTGATCTGCCGCCT 12425
|||||
RESULT 58
US-08-250-740-34
; Sequence 34, Application US/08250740
; Patent No. 5686240
; GENERAL INFORMATION:
; APPLICANT: Schuchman, Edward H.
; APPLICANT: Desnick, Robert J.
; TITLE OF INVENTION: Acid Sphingomyelinase Gene and Diagnosis
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/250,740
; FILING DATE: 27-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 6923-038
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1664 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-250-740-34
Query Match 0.2%; Score 46; DB 1; Length 1664;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9015 taattttgtatttttagtagatggggtttcaccatgttgcca 9060
|||||
Db 634 TAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTGGCCA 679
|||||
RESULT 59
US-07-695-472B-3
; Sequence 3, Application US/07695472B
; Patent No. 5773278
; GENERAL INFORMATION:

APPLICANT: Schuchman, Edward H.
APPLICANT: Desnick, Robert J.
TITLE OF INVENTION: The Acid Sphingomyelinase Gene and
TITLE OF INVENTION: Diagnosis of Niemann-Pick Disease
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/695,472B
FILING DATE: 19910503
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6923-014
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 7908864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1664 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-07-695-472B-3

Query Match 0.2%; Score 46; DB 1; Length 1664;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagatagatgggtttccaccatgttgcca 9060
|||||
Db 634 TAATTTTGTATTTTACTAGAGATGGGGTTTCACCATGTTGGCCA 679

RESULT 60
US-09-117-250-4
Sequence 4, Application US/09117250A
Patent No. 6251613
GENERAL INFORMATION:
APPLICANT: Kishimoto, Toshihiko
APPLICANT: Tamura, Taka-aki
APPLICANT: Makino, Yasutaka
TITLE OF INVENTION: METHOD OF DETECTING ANTI-DADII ANTIBODY AND METHOD OF
TITLE OF INVENTION: DIAGNOSING
TITLE OF INVENTION: CANCER USING THE DETECTING METHOD
FILE REFERENCE: 7898/255192
CURRENT APPLICATION NUMBER: US/09/117,250A
CURRENT FILING DATE: 1999-02-14
EARLIER APPLICATION NUMBER: JP8-11695
EARLIER FILING DATE: 1996-01-26
EARLIER APPLICATION NUMBER: PCT/JP97/00174
EARLIER FILING DATE: 1997-01-27
NUMBER OF SEQ ID NOS: 4
SEQ ID NO 4
LENGTH: 1926
TYPE: DNA
ORGANISM: Homo.sapiens
FEATURE:

NAME/KEY: CDS
LOCATION: (72)..(1553)
US-09-117-250-4

Query Match 0.2%; Score 46; DB 4; Length 1926;
Best Local Similarity 100.0%; Pred. No. 2.2e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8388 aggcgggtgcagtggtcagcgcctgttaatccacgactttgggag 8433
|||||
Db 1628 aggcgggtgcagtggtcagcgcctgttaatccacgactttgggag 1673

RESULT 61
US-08-687-080-55/C
Sequence 55, Application US/08687080
Patent No. 5965427
GENERAL INFORMATION:
APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 2886 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 5' END OF RAD50 GENOMIC SEQUENCE
US-08-687-080-55

Query Match 0.2%; Score 46; DB 2; Length 2886;
Best Local Similarity 100.0%; Pred. No. 2.1e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4709 gctaattttttagatagacgggtttccaccatgttg 4754
|||||
Db 174 GCTAATTTTGTATTTTACTAGACGGGGTTTCACCATGTTGG 129

RESULT 62
US-08-951-648-5/c

; Sequence 5, Application US/08951648
; Patent No. 5932465
; GENERAL INFORMATION:
; APPLICANT: Loughney, Kate
; TITLE OF INVENTION: Phosphodiesterase 8
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker, Sears Tower Suite 6300
; CITY: Chicago
; STATE: Illinois
; COUNTRY: US
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,648
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/34038
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 67..2403
; US-08-951-648-5

Query Match 0.2%; Score 46; DB 2; Length 3195;
Best Local Similarity 100.0%; Pred. No. 2.1e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17575 gacctcaggtgatccaccacccctcagctcccaaaagtgtgggatt 17620
|||||
DB 2627 GACCTCAGGTGATCCACCACCTCAGCTCCCAAAAGTGTGGGATT 2582

RESULT 63
US-09-174-437-5/C
; Sequence 5, Application US/09174437A
; Patent No. 6133007
; GENERAL INFORMATION:
; APPLICANT: Loughney, Kate
; TITLE OF INVENTION: Phosphodiesterase 8A
; FILE REFERENCE: 27866/35047
; CURRENT APPLICATION NUMBER: US/09/174,437A
; CURRENT FILING DATE: 1998-10-16
; EARLIER APPLICATION NUMBER: 08/951,648
; EARLIER FILING DATE: 1997-10-16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 3195
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (67)..(2403)
; US-09-174-437-5

Query Match 0.2%; Score 46; DB 3; Length 3195;
Best Local Similarity 100.0%; Pred. No. 2.1e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17575 gacctcaggtgatccaccacccctcagctcccaaaagtgtgggatt 17620
|||||
DB 2627 GACCTCAGGTGATCCACCACCTCAGCTCCCAAAAGTGTGGGATT 2582

RESULT 64
US-07-695-472B-4
; Sequence 4, Application US/07695472B
; Patent No. 5773278
; GENERAL INFORMATION:
; APPLICANT: Schuchman, Edward H.
; APPLICANT: Desnick, Robert J.
; TITLE OF INVENTION: The Acid Sphingomyelinase Gene and
; TITLE OF INVENTION: Diagnosis of Niemann-Pick Disease
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/695,472B
; FILING DATE: 19910503
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6923-014
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 7908864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4741 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-07-695-472B-4

Query Match 0.2%; Score 46; DB 1; Length 4741;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9015 taattttgtatttttagtagagatgggtttccaccatgttgcca 9060
|||||
DB 2324 TAATTTTGTATTATTAGTAGAGATGGGTTTCACCATGTGGCCA 2369

RESULT 65
US-08-250-740-35
; Sequence 35, Application US/08250740
; Patent No. 5686240
; GENERAL INFORMATION:
; APPLICANT: Schuchman, Edward H.
; APPLICANT: Desnick, Robert J.
; TITLE OF INVENTION: Acid Sphingomyelinase Gene and Diagnosis
; TITLE OF INVENTION: of Niemann-Pick Disease

NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/250,740
FILING DATE: 27-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30742
REFERENCE/DOCKET NUMBER: 6923-038
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 4742 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-250-740-35

Query Match 0.28; Score 46; DB 1; Length 4742;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagatagagtggtttccaccatgttgcca 9060
|||||
Db 2324 TAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGGCCA 2369

RESULT 66
US-08-471-058-20/c
Sequence 20, Application US/08471058
Patent No. 5770443
GENERAL INFORMATION:
APPLICANT: Kiefer, Michael C.
TITLE OF INVENTION: NOVEL APOPTOSIS MODULATING
TITLE OF INVENTION: PROTEINS, DNA ENCODING THE PROTEINS AND METHODS OF USE
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,058
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/320,157

FILING DATE: 07-OCT-1994
APPLICATION NUMBER: 08/160,067
FILING DATE: 30-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lehnhardt, Susan K.
REGISTRATION NUMBER: 33,943
REFERENCE/DOCKET NUMBER: 23647-20007.12
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-813-5600
TELEFAX: 415-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 5408 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1665...1928
OTHER INFORMATION:
US-08-471-058-20

Query Match 0.28; Score 46; DB 1; Length 5408;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9015 taattttgtatttttagatagagtggtttccaccatgttgcca 9060
|||||
Db 4880 TAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGGCCA 4835

RESULT 67
US-08-471-057-20/c
Sequence 20, Application US/08471057
Patent No. 6015887
GENERAL INFORMATION:
APPLICANT: KIEFER, MICHAEL C.
TITLE OF INVENTION: NOVEL APOPTOSIS-MODULATING PROTEINS, DNA
TITLE OF INVENTION: ENCODING THE PROTEINS AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,057
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/320,157
FILING DATE: 07-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: LEHNHARDT, SUSAN K.
REGISTRATION NUMBER: 33,943
REFERENCE/DOCKET NUMBER: 23647-20007.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 5408 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1665..1928
; US-08-471-057-20

Query Match 0.2%; Score 46; DB 3; Length 5408;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9015 taattttgtatttagtagagatgggtttccaccatgttgcca 9060
|||||
Db 4880 TAATTTTGTATTTTAGTAGAGATGGGTTTCACCATGTGGCCA 4835

RESULT 68

US-08-724-394A-20/c
; Sequence 20, Application US/08724394A
; Patent No. 5872237

GENERAL INFORMATION:

; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta

; APPLICANT: Wolff, Roger K.

; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP

; STREET: Two Embarcadero Center, 8th Floor

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/724,394A

; FILING DATE: 01-OCT-1996

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Fitts, Renee A.

; REGISTRATION NUMBER: 35,136

; REFERENCE/DOCKET NUMBER: 017957-000100

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-576-0200

; TELEFAX: 415-576-0300

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 246240 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 1..246240

; OTHER INFORMATION: /note= "HLA-H. CONTIG"

; US-08-724-394A-20

Query Match 0.2%; Score 46; DB 2; Length 246240;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8388 agccgggtgcagtgctcacgcctgtaatccccagcactttgggag 8433
|||||
Db 75431 AGCCGGGTGCAGTGGCTCAGCGCTGTAATCCCGACACTTTGGGAG 75386

RESULT 69

US-08-724-394A-21/c

; Sequence 21, Application US/08724394A

; Patent No. 5872237

; GENERAL INFORMATION:

; APPLICANT: Feder, John N.

; APPLICANT: Kronmal, Gregory S.

; APPLICANT: Lauer, Peter M.

; APPLICANT: Ruddy, David A.

; APPLICANT: Thomas, Winston

; APPLICANT: Tsuchihashi, Zenta

; APPLICANT: Wolff, Roger K.

; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el

; TITLE OF INVENTION: Sequences and Antibodies Thereto

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP

; STREET: Two Embarcadero Center, 8th Floor

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/724,394A

; FILING DATE: 01-OCT-1996

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Fitts, Renee A.

; REGISTRATION NUMBER: 35,136

; REFERENCE/DOCKET NUMBER: 017957-000100

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-576-0200

; TELEFAX: 415-576-0300

; INFORMATION FOR SEQ ID NO: 21:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 246240 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 1..246240

; OTHER INFORMATION: /note= "HLA-H. CONTIG"

; US-08-724-394A-21

Query Match 0.2%; Score 46; DB 2; Length 246240;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8388 agccgggtgcagtgctcacgcctgtaatccccagcactttgggag 8433
|||||
Db 75431 AGCCGGGTGCAGTGGCTCAGCGCTGTAATCCCGACACTTTGGGAG 75386

RESULT 70

US-08-724-394A-22/c

; Sequence 22, Application US/08724394A

; Patent No. 5872237

; GENERAL INFORMATION:

; APPLICANT: Feder, John N.

APPLICANT: Kronmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-22

Query Match 0.2%; Score 46; DB 2; Length 246240;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8388 aggcgggtgagtggtcagcctgtatccagcactttggag 8433
|||||
Db 75431 AGCGGGTGCGAGTGCTCAGCGCTGTATCCAGCACTTTGGGAG 75386

RESULT 71
US-08-687-080-76
Sequence 76, Application US/08687080
Patent No. 5965427
GENERAL INFORMATION:
APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 1386 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: INTRON 9 OF RAD50 GENOMIC-SEQUENCE
US-08-687-080-76

Query Match 0.2%; Score 45; DB 2; Length 1386;
Best Local Similarity 100.0%; Pred. No. 5.6e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19001 tgaaccccgctctactataaaatacaaaattagctggcggtg 19045
|||||
Db 453 TGAACCCCGCTCTACTAAAAATACAAAAATTAGCTGGCGTGG 497

RESULT 72
US-08-975-762-45
Sequence 45, Application US/08975762
Patent No. 6207169
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,762
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 45:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2373 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-975-762-45

Query Match 0.2%; Score 45; DB 4; Length 2373;
Best Local Similarity 100.0%; Pred. No. 5.3e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8389 ggcggggtgcagtggtcacgcctgtaatcccagcactttgggag 8433
|||||
Db 1972 GGCGGGTGCGAGTGCTCACGCTGTAATCCAGCACTTTGGGAG 2016

RESULT 73
US-08-171-382-5/c
; Sequence 5, Application US/08171382
; Patent No. 5472856
; GENERAL INFORMATION:
; APPLICANT: Harris, Crafford A.
; APPLICANT: Goldstein, Gideon
; APPLICANT: Siekierka, John J.
; APPLICANT: Talle, Mary Anne
; APPLICANT: Shenbagamurthi, Ponniah
; APPLICANT: Culler, Michael D.
; APPLICANT: Setcavage, Diane R.
; TITLE OF INVENTION: Recombinant Human Thymopoietin Proteins
; TITLE OF INVENTION: and Uses Therefor
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/171,382
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: IRI43USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2392 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 241..1275
US-08-171-382-5

Query Match 0.2%; Score 45; DB 1; Length 2392;
Best Local Similarity 100.0%; Pred. No. 5.2e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10503 ttttttagtagagacggggtttccaccatgttgccaggatggtc 10547
|||||
Db 2383 TTTTtagtagagacggggtttccaccatgttgccaggatggtc 2339

RESULT 74
US-08-309-420-5/c
; Sequence 5, Application US/08309420
; Patent No. 5591588
; GENERAL INFORMATION:
; APPLICANT: Goldstein, Gideon
; APPLICANT: Culler, Michael
; TITLE OF INVENTION: Method for the Diagnosis of Depression
; TITLE OF INVENTION: Based on Monitoring Blood Levels of Arginine Vasopressin
; TITLE OF INVENTION: and/or Thymopoietin
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: 321 No. 5591588ristown Road, Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,420
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: IRI46USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 540-9207
; TELEFAX: (215) 540-5818

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2392 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 241..1275
US-08-309-420-5

Query Match 0.2%; Score 45; DB 1; Length 2392;
Best Local Similarity 100.0%; Pred. No. 5.2e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10503 ttttttagtagagacggggtttccaccatgttgccaggatggtc 10547
|||||
Db 2383 TTTTtagtagagacggggtttccaccatgttgccaggatggtc 2339

RESULT 75
US-08-309-419-5/c
; Sequence 5, Application US/08309419
; Patent No. 5593842
; GENERAL INFORMATION:
; APPLICANT: Goldstein, Gideon
; APPLICANT: Culler, Michael
; TITLE OF INVENTION: Method of Measuring Thymopoietin
; TITLE OF INVENTION: Proteins in Plasma and Serum
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson

STREET: 321 No. 5593842ristown Road, Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: US/08/309,419
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: IRI45BUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 540-9206
TELEFAX: (215) 540-5818
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 241..1275
US-08-309-419-5

Query Match 0.2%; Score 45; DB 1; Length 2392;
Best Local Similarity 100.0%; Pred. No. 5.2e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10503 ttttttagtagagacggggtttaccatgttgccaggatggtc 10547
|||||
Db 2383 TTTTtTAGTAGAGACGGGGTTTcACCATGTTGCCAGGATGGTC 2339

RESULT 76
PCT-US95-11856-5/c
Sequence 5, Application PC/TUS9511856
GENERAL INFORMATION:
APPLICANT: Immunobiology Research, Institute Inc.
TITLE OF INVENTION: Method of Measuring
TITLE OF INVENTION: Thymopoietin Proteins in Plasma and Serum
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: 321 Norristown Road, Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version#1.25
CURRENT APPLICATION DATA:
FILING DATE: PCT/US95/11856
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/309,419
FILING DATE: 20-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: IRI45BPCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 540-9206
TELEFAX: (215) 540-5818
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 241..1275
PCT-US95-11856-5

Query Match 0.2%; Score 45; DB 5; Length 2392;
Best Local Similarity 100.0%; Pred. No. 5.2e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10503 ttttttagtagagacggggtttaccatgttgccaggatggtc 10547
|||||
Db 2383 TTTTtTAGTAGAGACGGGGTTTcACCATGTTGCCAGGATGGTC 2339

RESULT 77
PCT-US95-11878-5/C
Sequence 5, Application PC/TUS9511878
GENERAL INFORMATION:
APPLICANT: Immunobiology Research, Institute Inc.
TITLE OF INVENTION: Method for the Diagnosis of
TITLE OF INVENTION: Depression Based on Monitoring Blood Levels of
TITLE OF INVENTION: Arginine Vasopressin and/or Thymopoietin
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: 321 Norristown Road, Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version#1.25
CURRENT APPLICATION DATA:
FILING DATE: PCT/US95/11878
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/309,420
FILING DATE: 20-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: IRI46PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 540-9206
TELEFAX: (215) 540-5818
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 241..1275
PCT-US95-11878-5

Query Match 0.2%; Score 45; DB 5; Length 2392;
Best Local Similarity 100.0%; Pred. No. 5.2e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10503 ttttttagtagagacgggtttccaccatgttgccaggatggtc 10547
|||||
Db 2383 TTTTGTAGTAGACGGGTTTCCACCATGTTGCCAGGATGTC 2339

RESULT 78

US-09-085-199B-44/c
; Sequence 44, Application US/09085199B
; Patent No. 6235879
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael R.
; APPLICANT: Hackam, Abigail
; APPLICANT: Huq, A.H.M. Mahbubul
; APPLICANT: Chopra, Vikramjit Singh
; APPLICANT: Kalchman, Michael
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the
; HUNTINGTON'S DISEASE GENE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppedahl & Larson
; STREET: PO Box 5270
; CITY: Frisco
; STATE: CO
; COUNTRY: USA
; ZIP: 80443-5270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS DOS 5.0
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,199B
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Larson, Marina T.
; REGISTRATION NUMBER: 32038
; REFERENCE/DOCKET NUMBER: UBC.P-013US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (970) 668-2050
; TELEFAX: (970) 668-2052
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3715
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: human
; FEATURE:
; OTHER INFORMATION: exon 29 and partial cds of HPI1
US-09-085-199B-44

Query Match 0.2%; Score 45; DB 4; Length 3715;
Best Local Similarity 100.0%; Pred. No. 5e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4715 ttttttagtagagacgggtttccaccatgttgccagg 4759
|||||
Db 3651 TTTTGTATTTTAGTAGACGGGTTTCCACCATGTTGCCAGG 3607

RESULT 79

US-08-757-223-7/c

; Sequence 7, Application US/08757223
; Patent No. 6136530
; GENERAL INFORMATION:
; APPLICANT: Poduslo, Shirley E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ASSESSING RISK
; FACTORS IN ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Locke Purnell Rain Harrell
; STREET: 2200 Ross Avenue, Suite 2200
; CITY: Dallas
; STATE: Texas
; ZIP: 75201-6776
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,223
; FILING DATE: No. 6136530ember 27, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mayfield, Denise L.
; REFERENCE/DOCKET NUMBER: 4-003US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 214/740-8785
; TELEFAX: 214/740-8800
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5375 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-757-223-7

Query Match 0.2%; Score 45; DB 3; Length 5375;
Best Local Similarity 100.0%; Pred. No. 4.8e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4588 ccaggctggagtcgagtcgcatgtctcgctcactgcacactcc 4632
|||||
Db 3754 CCAGGCTGGAGTCGAGTGGCATGATCTCGGCTCACTGCAACCTCC 3710

RESULT 80

US-08-611-587-1/c
; Sequence 1, Application US/08611587
; Patent No. 6150091
; GENERAL INFORMATION:
; APPLICANT: PANDOLFO, MASSIMO
; APPLICANT: MONTERMINI, LAURA
; APPLICANT: MOLTO, MARIA D.
; APPLICANT: Koenig, Michael
; APPLICANT: Campuzano, Victoria
; APPLICANT: Cossee, Mireille
; TITLE OF INVENTION: Direct Diagnosis of Friedreich Ataxia
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P. Patent Dept.
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.
; ZIP: 77010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611,587

; FILING DATE: 03-MAR-1996
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Brashears-Macatee, Sarah J.
; REGISTRATION NUMBER: 38,087
; REFERENCE/DOCKET NUMBER: D-5901
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-651-5620
; TELEFAX: 713-651-5246
; TELEX: 76-2829
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8353 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 9q13
; UNITS: bp
; US-08-611-587-1

Query Match 0.2%; Score 45; DB 3; Length 8353;
Best Local Similarity 100.0%; Pred. No. 4.6e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4715 ttttggatatttagtagagcgggtttccaccatgttggccagg 4759
|||||
Db 6281 TTTTGTATTTTAGTAGAGCGGGTTTCACCATGTTGGCCAGG 6237

RESULT 81
US-08-750-703-4/c
; Sequence 4, Application US/08750703
; Patent No. 5891633
; GENERAL INFORMATION:
; APPLICANT: Gonzalez, Frank J.; Idle, Jeffrey R.
; TITLE OF INVENTION: DEFECTS IN DRUG
; TITLE OF INVENTION: METABOLISM
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan
; STREET: 345 Park Ave.
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,703
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07605
; FILING DATE: 16-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Dorothy R. Auth
; REGISTRATION NUMBER: 36,434
; REFERENCE/DOCKET NUMBER: 2026-4196PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 8779 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: CYP2A13
; LOCATION:
; OTHER INFORMATION:
; US-08-750-703-4

Query Match 0.2%; Score 45; DB 2; Length 8779;
Best Local Similarity 100.0%; Pred. No. 4.5e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4715 ttttggatatttagtagagcgggtttccaccatgttggccagg 4759
|||||
Db 4352 TTTTGTATTTTAGTAGAGCGGGTTTCACCATGTTGGCCAGG 4308

RESULT 82
US-08-484-044-10/c
; Sequence 10, Application US/08484044
; Patent No. 5552282
; GENERAL INFORMATION:
; APPLICANT: Caskey, C. T.
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Friedman, David L.
; APPLICANT: Pizzuti, Antonio
; APPLICANT: Fenwick, Raymond G.
; TITLE OF INVENTION: Diagnosis of Myotonic Muscular Dystrophy
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski, L.L.P.
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,044
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/019,940
; FILING DATE: 19-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5325
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11613 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-484-044-10

Query Match 0.2%; Score 45; DB 1; Length 11613;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17555 cagcgtggtctcgaactcctgacctcaggtgagtcacccacacctca 17599
|||||
Db 6776 CAGCGTGGTCTCGAAGCTCCTGACCTCAGGTGATCCACCCACCTCA 6732

RESULT 83

US-08-076-011-1/c
; Sequence 1, Application US/08076011
; Patent No. 5521069
; GENERAL INFORMATION:
; APPLICANT: ONDA, Haruo
; APPLICANT: KIMURA, Chiharu
; APPLICANT: OKUBO, Shoichi
; TITLE OF INVENTION: NOVEL DNA AND USE THEREOF
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN,
; ADDRESSEE: ROBERTS & CUSHMAN
; STREET: 130 Water Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,011
; FILING DATE: 11-JUN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/047,246
; FILING DATE: 13-APR-1993
; NAME:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,676
; FILING DATE: 07-AUG-1991
; NAME:
; ATTORNEY/AGENT INFORMATION:
; NAME: RESNICK, DAVID S
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 41155-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)523-3400
; TELEFAX: (617)523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(7540..7650, 9814..9945, 10421..10519,
; LOCATION: 11602..11787)
US-08-076-011-1

Query Match 0.2%; Score 45; DB 1; Length 17041;
Best Local Similarity 100.0%; Pred. No. 4.2e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9018 ttttgtatttttagtagagatgggtttcaccatgttgccagg 9062
|||||
Db 73 TTTTGTATTTTATGATGATGGGTTCACCATGTGTCAGG 29

RESULT 84

US-09-318-448-11
; Sequence 11, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Steenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 18596
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-11

Query Match 0.2%; Score 45; DB 4; Length 18596;
Best Local Similarity 100.0%; Pred. No. 4.2e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8932 cctccacctcccaggttcaagtattctcctgcctcagctccca 8976
|||||
Db 17353 cctccacctcccaggttcaagtattctcctgcctcagctccca 17397

RESULT 85

US-08-814-095-7
; Sequence 7, Application US/08814095
; Patent No. 6025183
; GENERAL INFORMATION:
; APPLICANT: Soreq, Hermona
; APPLICANT: Zakut, Haim
; APPLICANT: Shani, Moshe
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
; TITLE OF INVENTION: ANTI-CHOLINESTERASE SUBSTANCES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KOHN & ASSOCIATES
; STREET: 30500 No. 6025183thwestern Highway, Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814,095
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Montgomery, Ilene N.
; REGISTRATION NUMBER: 38,972
; REFERENCE/DOCKET NUMBER: 2391.00066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (248) 539-5050
; TELEFAX: (248) 539-5055
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "cosmid including ACHE
; DESCRIPTION: promotor, ACHE gene and ARS gene"
; HYPOTHETICAL: NO

```
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 7q22
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 4089..22464
; OTHER INFORMATION: /function= "ACHE Promotor"
; OTHER INFORMATION: /standard_name= "ACHE Promotor"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 22465..22537
; OTHER INFORMATION: /function= "non-translated"
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: exon
; LOCATION: 24090..25177
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /function= "(translation start:
; OTHER INFORMATION: 24110)"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25524..26009
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 27005..27274
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 27255..28007
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 5
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 27385..27387
; FEATURE:
; NAME/KEY: exon
; LOCATION: 28008..28129
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 28129..28131
; FEATURE:
; NAME/KEY: exon
; LOCATION: 34528..34895)
; OTHER INFORMATION: /function= "arsenite resistance
; OTHER INFORMATION: gene"
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: exon
; LOCATION: 34092..34358)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 2
; FEATURE:
; NAME/KEY: exon
```

```
; LOCATION: complement (33779..33963)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 3
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (33493..33591)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 4
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (33297..33408)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 5
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (32959..33094)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (32569..32628)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 7
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (32386..32468)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 8
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (31894..32080)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 9
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (31363..31534)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 10
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (31131..31284)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 11
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (30816..31011)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 12
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (30470..30626)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 13
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (30187..30274)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 14
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (29945..30073)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 15
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (29664..29856)
; OTHER INFORMATION: /gene= "ARS"
; OTHER INFORMATION: /number= 16
; US-08-814-095-7
```

Query Match 0.2%; Score 45; DB 3; Length 35060;
Best Local Similarity 100.0%; Pred. No. 3.9e-08;

```
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8389 ggcgggtcagtgctcacgctgtaatccagcacttggag 8433
|||||
Db 6202 GCGCGGTGAGTGGCTCACGCTGTATCCAGCACTTTGGGAG 6246
|||||

RESULT 86
US-08-781-891-79
; Sequence 79, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 87350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-79

Query Match 0.2%; Score 45; DB 3; Length 87350;
Best Local Similarity 100.0%; Pred. No. 3.5e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10571 gctgccttgacctccaaagtctggattacaggctgagcca 10615
|||||
Db 72341 GCGTCCCTTGGCTCCCAAGTCTGGGATTACAGCGGTGAGCCA 72385
|||||

RESULT 87
US-08-222-177A-10/C
; Sequence 10, Application US/08222177A
; Patent No. 5582979
; GENERAL INFORMATION:
; APPLICANT: Weber, James L.
; TITLE OF INVENTION: LENGTH POLYMORPHISMS IN
; TITLE OF INVENTION: (dC-dA)n.(dG-dT)n SEQUENCES AND METHODS OF USING SAME
; NUMBER OF SEQUENCES: 460
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dewitt Ross & Stevens, S.C.
; STREET: 8000 Excelsior Drive, Suite 401
```

```
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53717-1914
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,177A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/341,562
; FILING DATE: 21-APR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Sara, Charles S.
; REGISTRATION NUMBER: 30,492
; REFERENCE/DOCKET NUMBER: 09865.601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 831-2100
; TELEFAX: (608) 831-2106
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 264 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; INDIVIDUAL ISOLATE: Caucasian
; TISSUE TYPE: Blood
; IMMEDIATE SOURCE:
; CLONE: Mfd22
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 4
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 153..193
; OTHER INFORMATION: /rpt_type= "tandem"
; OTHER INFORMATION: /rpt_family= "(dC-dA)n.(dG-dT)n"
; OTHER INFORMATION: /citation= ([2])
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 89..107
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "PCR primer"
; OTHER INFORMATION: /citation= ([1])
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: complement (220..238)
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "PCR primer"
; OTHER INFORMATION: /citation= ([1])
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..264
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "Only one strand sequenced"
; PUBLICATION INFORMATION:
; AUTHORS: Weber, J. L.
; AUTHORS: May, P. E.
; TITLE: Dinucleotide repeat polymorphism at the
; TITLE: D4S171 locus
; JOURNAL: Nucleic Acids Res.
; VOLUME: 18
; PAGES: 2202-
```

DATE: 1990
PUBLICATION INFORMATION:
AUTHORS: Weber, James L.
AUTHORS: May, Paula E.
TITLE: Abundant Class of Human DNA Polymorphisms
TITLE: Which Can Be Typed Using the Polymerase Chain
TITLE: Reaction
JOURNAL: Am. J. Hum. Genet.
VOLUME: 44
PAGES: 388-396
DATE: 1989
US-08-222-177A-10

Query Match 0.2%; Score 44; DB 1; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10369 ccaggctgagtgagtgatgcttctgctcaactgcaactc 10412
|||||
Db 91 CGAGCTGGAGTGCAGTGTGGATCTTGGCTCACTGCAACCTC 48

RESULT 88
US-08-481-658B-63/c
Sequence 63, Application US/08481658B
Patent No. 5955075
GENERAL INFORMATION:

APPLICANT: Zavada, Jan
APPLICANT: Pastorekova, Silvia
APPLICANT: Pastorek, Jaromir
TITLE OF INVENTION: MN Gene and Protein
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leona L. Lauder
STREET: 6 Mariposa Court
CITY: Tiburon
STATE: California
COUNTRY: USA
ZIP: 94920

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481.658B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/260,190
FILING DATE: 15-JUN-1994
ATTORNEY/AGENT INFORMATION:

NAME: Lauder, Leona L.
REGISTRATION NUMBER: 30,863
REFERENCE/DOCKET NUMBER: D-0021.3E
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-435-2034
TELEFAX: 415-435-0727

INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-481-658B-63

Query Match 0.2%; Score 44; DB 2; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10572 cctgccttggcctcccaagtcgtggattacagcgtgagcca 10615
|||||
Db 56 CTTGCTTGGCTCCCAAGTCTGGGATTACAGCGGTGAGCCA 13

RESULT 89
US-08-477-504A-63/c
Sequence 63, Application US/08477504A
Patent No. 5972353
GENERAL INFORMATION:
APPLICANT: Zavada, Jan
APPLICANT: Pastorekova, Silvia
APPLICANT: Pastorek, Jaromir
TITLE OF INVENTION: MN Gene and Protein
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leona L. Lauder
STREET: 6 Mariposa Court
CITY: Tiburon
STATE: California
COUNTRY: USA
ZIP: 94920

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477.504A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/260,190
FILING DATE: 15-JUN-1994
ATTORNEY/AGENT INFORMATION:

NAME: Lauder, Leona L.
REGISTRATION NUMBER: 30,863
REFERENCE/DOCKET NUMBER: D-0021.3D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-435-2034
TELEFAX: 415-435-0727

INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-477-504A-63

Query Match 0.2%; Score 44; DB 2; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10572 cctgccttggcctcccaagtcgtggattacagcgtgagcca 10615
|||||
Db 56 CTTGCTTGGCTCCCAAGTCTGGGATTACAGCGGTGAGCCA 13

RESULT 90
US-08-486-756A-63/c
Sequence 63, Application US/08486756A
Patent No. 5981711
GENERAL INFORMATION:
APPLICANT: Zavada, Jan
APPLICANT: Pastorekova, Silvia
APPLICANT: Pastorek, Jaromir
TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,756A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; US-08-486-756A-63

Query Match 0.2%; Score 44; DB 2; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10572 cctgccttgccctcccaagtcgtggattacagcgtagacca 10615
|||||
Db 56 CCGCTTGCCCTCCCAAGTCTGGGATTACAGCGTGAGCCA 13

RESULT 91
US-08-485-862B-63/C
; Sequence 63, Application US/08/485862B
; Patent No. 5989838
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,862B
; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEtical: NO
; ANTI-SENSE: NO
; US-08-485-862B-63

Query Match 0.2%; Score 44; DB 2; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10572 cctgccttgccctcccaagtcgtggattacagcgtagacca 10615
|||||
Db 56 CCGCTTGCCCTCCCAAGTCTGGGATTACAGCGTGAGCCA 13

RESULT 92
US-08-787-739-63/C
; Sequence 63, Application US/08/787739
; Patent No. 6027887
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 369 Pine Street, Suite 610
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/787,739
; FILING DATE: 24-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,049
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,756
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/481,658
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,862

; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,863
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,077
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-981-2034
; TELEFAX: 415-981-0332
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-787-739-63

Query Match 0.2%; Score 44; DB 3; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10572 cctgccttggcctcccaagtgctgggattacagcggtgagcca 10615
|||||
DB 56 CTTGCTTGGCTCCCAAGTCTGGGATTACAGCGGTGAGCCA 13

RESULT 93
US-08-487-077A-63/c
; Sequence 63, Application US/08487077A
; Patent No. 6069242
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,077A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3H
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-077A-63

Query Match 0.2%; Score 44; DB 3; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10572 cctgccttggcctcccaagtgctgggattacagcggtgagcca 10615
|||||
DB 56 CTTGCTTGGCTCCCAAGTCTGGGATTACAGCGGTGAGCCA 13

RESULT 94
US-08-485-863A-63/c
; Sequence 63, Application US/08485863A
; Patent No. 6093548
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,863A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-485-863A-63

Query Match 0.2%; Score 44; DB 3; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10572 cctgccttggcctcccaagtgctgggattacagcggtgagcca 10615
|||||

Db 56 CcTgcCTTGGcCTCCAAAGTGTGGATTACAGCGGTGAGCCA 13

RESULT 95

US-08-485-049D-63/c
; Sequence 63, Application US/08485049D
; Patent No. 6204370
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 369 Pine Street
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,049D
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-981-2034
; TELEFAX: 415-981-0332
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-485-049D-63

Query Match 0.2%; Score 44; DB 4; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10572 cctgccttgccctccaaagtgtggattacagcggtgagcca 10615
|||||

Db 56 CcTgcCTTGGcCTCCAAAGTGTGGATTACAGCGGTGAGCCA 13

RESULT 96

US-08-481-658B-44/c
; Sequence 44, Application US/08481658B
; Patent No. 5955075
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court

; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,658B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; DESCRIPTION: 6th MN intron
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-481-658B-44

Query Match 0.2%; Score 44; DB 2; Length 1334;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10572 cctgccttgccctccaaagtgtggattacagcggtgagcca 10615
|||||

Db 264 CcTgcCTTGGcCTCCAAAGTGTGGATTACAGCGGTGAGCCA 221

RESULT 97

US-08-477-504A-44/c
; Sequence 44, Application US/08477504A
; Patent No. 5972353
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,504A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; DESCRIPTION: 6th MN intron
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-477-504A-44

Query Match 0.2%; Score 44; DB 2; Length 1334;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10572 cctgccttggcctcccaaaagtctgggattacagcggtgagcca 10615
|||||
Db 264 CTGCTTGGCCTCCCAAAGTCTGGGATTACAGCGGTGAGCCA 221

RESULT 98
US-08-486-756A-44/c
; Sequence 44, Application US/08486756A
; Patent No. 5981711

; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,756A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

; DESCRIPTION: 6th MN intron
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-486-756A-44

Query Match 0.2%; Score 44; DB 2; Length 1334;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10572 cctgccttggcctcccaaaagtctgggattacagcggtgagcca 10615
|||||
Db 264 CTGCTTGGCCTCCCAAAGTCTGGGATTACAGCGGTGAGCCA 221

RESULT 99
US-08-485-862B-44/c
; Sequence 44, Application US/08485862B
; Patent No. 5989838

; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,862B
; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995

; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1334 base pairs
; TYPE: nucleic acid

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

; DESCRIPTION: 6th MN intron
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-485-862B-44

Query Match 0.2%; Score 44; DB 2; Length 1334;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10572 cctgccttggcctcccaaaagtctgggattacagcggtgagcca 10615
|||||
Db 264 CTGCTTGGCCTCCCAAAGTCTGGGATTACAGCGGTGAGCCA 221

RESULT 100

US-08-787-739-44/c
; Sequence 44, Application US/08787739
; Patent No. 6027887
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 369 Pine Street, Suite 610
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/787,739
; FILING DATE: 24-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,049
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,756
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/481,658
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,862
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,863
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,077
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-981-2034
; TELEFAX: 415-981-0332
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; DESCRIPTION: 6th MN intron
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-787-739-44

Query Match 0.2%; Score 44; DB 3; Length 1334;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10572 cctgccttggcctcccaagtgctgggttacagcggtgagcca 10615

Db 264 CCTGCCTTGGCCTCCCAAGTGGCTGGGATTACAGCGTGAGCCA 221
|||||

Search completed: November 3, 2001, 00:58:55
Job time: 99845 sec

